# Inferência Causal em Σpidemiologia

## o modelo de respostas potenciais

# Inferência Causal em Σpidemiologia

## o modelo de respostas potenciais

Ronir Raggio Luiz
Claudio José Struchiner

ISBN: 85-7541-010-5

Capa, projeto gráfico e editoração eletrônica
*Angélica Mello*

Preparação de originais e revisão
*Marcionílio Cavalcanti de Paiva*

Copidesque
*Irene Ernest Dias*

Catalogação-na-fonte
Centro de Informação Científica e Tecnológica
Biblioteca Lincoln de Freitas Filho

---

L954i Luiz, Ronir Raggio
Inferência causal em epidemiologia: o modelo de respostas potenciais. / Ronir Raggio Luiz e Claudio José Struchiner. Rio de Janeiro: Editora Fiocruz, 2002.
112 p., tab.

1. Inferência. 2. Causalidade. 3. Modelos epidemiológicos.

CDD – 20.ed. – 614.4

---

2002
EDITORA FIOCRUZ
Av. Brasil, 4036 – 1º andar – sala 112 – Manguinhos
21040-361 – Rio de Janeiro – RJ
Tels.: (21) 3882-9039 e 3882-9041
Telefax: (21) 3882-9006 e 3882-9007
http://www.fiocruz.br/editora
e-mail: editora@fiocruz.br

# Sumário

# Prefácio

É comum caracterizar a epidemiologia como uma disciplina (para alguns) ou ciência (para outros) que tem como um dos seus eixos centrais de interesse a preocupação com a determinação do processo saúde-doença. Assim, é compreensível que a causalidade seja um tema recorrente na literatura produzida na área. A produção intelectual sobre esse assunto, dada sua complexidade, envolve uma multiplicidade de abordagens, configurando uma rede de argumentos e proposições que exploram desde aspectos mais filosóficos – como a construção e a crítica do determinismo causal –, até questões mais operacionais – como os critérios de julgamento da causalidade –, passando por discussões sobre os limites e as possibilidades de sua aproximação por meio do método científico. Explorar toda essa diversidade de pontos de vista é uma tentação para quem deseja ter uma compreensão mais global do problema. Afinal, como seria possível falar de causalidade sem se reportar aos pensamentos originais de Aristóteles, ao cartesianismo, ao 'problema de Hume', ao critério de demarcação de ciência delineado por Popper e aos tipos de determinação propostos por Bunge? Como refletir sobre causalidade em epidemiologia sem conhecer os postulados de Henle-Koch, os critérios de Hill, as idéias de Susser e o modelo de causas suficiente, necessária e componente de Rothman? Como discutir inferência causal na epidemiologia moderna sem considerar os pensamentos de Miettinen, Robins

e Greenland sobre confundimento? E como ignorar os desenvolvimentos mais recentes sobre diagramas causais, propostos por Pearl? Entretanto, se a multiplicidade de vertentes é a regra, ela leva, muitas vezes, a um discurso demasiadamente genérico. A grata surpresa que o leitor terá com este livro é a objetividade dos autores, que, evitando falar sobre 'tudo ao mesmo tempo agora', propõem um mergulho elucidativo em uma das mais importantes contribuições da estatística para a discussão da causalidade, nomeadamente o modelo de respostas potenciais de Rubin, e sua interface com o método epidemiológico.

A base contrafactual do modelo de respostas potenciais de Rubin provoca uma certa inquietação nos epidemiologistas, pois implica imediatamente uma referência ao problema fundamental da inferência causal, qual seja, a impossibilidade de se observar respostas para diferentes tratamentos em um mesmo indivíduo, simultaneamente. Nessa perspectiva, conforme salientam os autores, o efeito causal em um só indivíduo poderia ser aferido se soubéssemos qual seria seu desfecho se ele tivesse sido exposto à causa potencial *t* e se seu desfecho tivesse sido exposto à causa potencial *c*. Como todo efeito de uma determinada causa *t* deve ser mensurado em relação a uma causa *c*, torna-se, então, impossível observar o efeito de *t* relativo a *c* naquele indivíduo. Os autores apresentam algumas possibilidades de superação desse impasse e discutem as condições para sua utilização epidemiológica. De particular interesse são as soluções científica e estatística, e as premissas adjacentes de homogeneidade e estabilidade. Dentre as formas mais comuns de ruptura da hipótese de estabilidade, os autores ressaltam a existência de interferência entre as unidades de observação, situação bastante comum em doenças transmissíveis. Uma outra questão fundamental para a formulação estatística de causalidade é a independência entre as respostas e o mecanismo de designação de tratamento. Os autores abordam esse problema sob a situação clássica de alocação

aleatória do tratamento nas unidades de observação (randomização) e sob a ótica dos estudos observacionais, em que se espera que tal independência esteja condicionada a um conjunto de co-variáveis observadas. Nesse momento, a idéia de um escore de propensão é apresentada de forma clara e heurística, acessível aos leitores com menor experiência estatística. Um pouco mais complexa, mas nem por isso menos importante ou menos interessante, é a seção sobre inferência estatística, em que os autores discutem quatro métodos conceitualmente distintos de inferência estatística: os testes de hipóteses nulas pontuais sob randomização, a inferência de parâmetros causais baseada em possíveis designações aleatórias do tratamento, a inferência bayesiana para efeitos causais e a inferência freqüentista. No capítulo seguinte, o leitor irá encarar, sob uma nova ótica, conceitos mais familiares ao epidemiologista. Nesse ponto, discutem-se fundamentalmente questões relacionadas ao julgamento da validade em estudos epidemiológicos, com particular ênfase nos conceitos de confundimento, permutabilidade e ignorabilidade, e as implicações para os tipos de delineamentos mais utilizados em epidemiologia. Finalmente, os autores apresentam um excelente texto de conclusão, mas não vale começar por ele!

Há muito era desejado um livro sobre inferência causal que apresentasse a contribuição da estatística para a discussão da causalidade, explorando suas interfaces com o método epidemiológico. O trabalho aqui apresentado cumpre esse objetivo, estabelecendo um diálogo com a epidemiologia em que são valorizadas e resguardadas as identidades de cada disciplina. Assim como outras disciplinas da área da saúde coletiva devem procurar ampliar seus horizontes e interagir com diferentes áreas do conhecimento, tanto a (bio)estatística como a epidemiologia têm muito a ganhar com o estreitamento de suas relações. É esse espírito interdisciplinar que torna o trabalho de Ronir e Claudio uma referência

importante, não só para os epidemiologistas e bioestatísticos, mas também para outros profissionais da saúde coletiva que desejem ampliar sua visão sobre a questão da causalidade.

*Guilherme Loureiro Werneck*
Núcleo de Estudos de Saúde Coletiva/
Depto. de Medicina Preventiva da UFRJ
Instituto de Medicina Social da UERJ

# Apresentação

Este livro tem sua origem na dissertação de mestrado em estatística desenvolvida por Ronir Raggio Luiz no Instituto de Matemática da Universidade Federal do Rio de Janeiro, tendo surgido do fascínio comum entre orientado e orientador pelos desafios da pesquisa biomédica e da epidemiologia. Ambos vislumbramos na inferência causal um tema que respondesse às nossas inquietações sobre a lógica da investigação científica.

A proposta, aqui, é apresentar a contribuição da estatística para a discussão sobre causalidade, por meio do modelo de respostas potenciais proposto por Rubin. Apesar da aridez do tema, procuramos traduzir para o leitor interessado, tipicamente um epidemiologista, questões presentes apenas em artigos científicos, e relacionadas operacional e conceitualmente a uma investigação causal. Nosso objetivo é oferecer ao pesquisador da área médica uma ponte entre a estatística e os aspectos causais da investigação epidemiológica. Se, de um lado, um estatístico tradicionalmente se esquiva quando confrontado com uma questão causal, de outro, um epidemiologista não tem contato com a noção estatística de causa. A palavra 'traduzir' se aplica também em seu sentido original, na medida em que, até aonde vai nosso conhecimento, não há textos em português sobre o tema.

Este volume está estruturado em três capítulos. Na Introdução, convidamos o leitor a se envolver com o tema. No capítulo 1 oferecemos um panorama da discussão de causalidade em epidemiologia e no capítulo 2

desenvolvemos a teoria propriamente dita do modelo de respostas potenciais, explorando o papel do mecanismo de designação de tratamento. O tópico intitulado Inferência Estatística é um pouco mais árido, mas sua leitura pode ser deixada de lado num primeiro momento sem comprometimento da essência do assunto. No capítulo 3 procuramos relacionar a teoria exposta a conceitos epidemiológicos, particularmente os de validade e confundimento. Por fim, na Conclusão discutimos as principais implicações do conteúdo apresentado. Em síntese, o modelo causal apresentado é um modelo construído sob uma ótica contrafactual, em que a variável resposta é, na realidade, um vetor de respostas potenciais, com dimensão dada pelo número de tratamentos considerados. A lógica subjacente a esse modelo é a busca dos efeitos de causas postuladas, e não o inverso, isto é, a busca das causas de efeitos observados. Essa mudança de caminho, apesar de revolucionária, não é nova nem originariamente própria à estatística. Conta-se que Albert Einstein, quando professor da Politécnica de Zurique, causou verdadeiro escândalo entre seus colegas ao afirmar que o princípio básico de toda a ciência superior era priori-dedutivo, e não posteriori-indutivo. Em outras palavras, o homem deve focalizar a 'causa' e daí partir para os 'efeitos'.

Um ponto que merece destaque quanto ao conteúdo do livro é a superação da até então incômoda interface da estatística com a noção de causa. Assim, sustentados na retórica estatística segundo a qual correlação não implica causalidade, e apesar do reconhecimento geral de que um experimento randomizado bem planejado pode fornecer poderosa ajuda na investigação de relações causais, os estatísticos abstiveram-se de pensar em causalidade de forma mais específica e audaciosa. Esse fato é sutilmente evidenciado por Oscar Kempthorne, um dos maiores estatísticos do século XX, recentemente falecido, quando apresenta a questão do que seja um conceito viável de causa. A substituição do foco da questão causal de

uma definição epistemológica para a mensuração de efeitos causais parece ser uma importante estratégia que os estatísticos encontraram para prestar a sua contribuição ao tema.

Por fim, vale a pena enfatizar a oportunidade do tema proposto, que deve ir além dos interesses da epidemiologia e da estatística, para ser percebido também como estratégico para o desenvolvimento da área da saúde pública de forma mais geral.

Os Autores

# Introdução

A curiosidade e a necessidade há muito têm estimulado a busca das causas dos diversos fenômenos que são rotineiramente observados. Em particular, a ocorrência de doenças é um fenômeno em que há interesse geral na identificação de suas causas para que obviamente possam ser prevenidas. Entretanto, inferir causalidade é uma tarefa complexa, envolvendo diversas áreas de investigação. A filosofia, a sociologia e a medicina sempre se sentiram desafiadas por essa questão, enquanto a estatística só mais recentemente parece ter despertado seu interesse por ela. Embora se identifiquem referências à idéia de causa nos trabalhos sobre experimentos aleatorizados desenvolvidos por Fisher no início do século XX, formalmente a contribuição da estatística para esta discussão começou com o trabalho de Rubin em 1974.

Holland (1986) salientou que quando se fala de causalidade, a dificuldade está na diversidade de questões que surgem. Os filósofos, por exemplo, estão interessados no significado fundamental da noção de causa. Os sociólogos ou médicos, por sua vez, interessam-se pela identificação das causas de um dado efeito. E há ainda os cientistas em geral, interessados em entender os detalhes dos mecanismos causais.

A discussão sobre causalidade parece ser mais objetiva em um contexto estatístico, uma vez que sua contribuição se concentra principalmente na mensuração de efeitos causais. Medir efeitos causais sem o entendimento

do mecanismo causal envolvido ou do significado de causalidade não só é possível como faz parte do cotidiano de todos. Eventualmente, pode ser mais conveniente em um primeiro momento tentar objetivamente, por meio de mensurações adequadas, identificar a causa de efeitos observados do que conceituá-la formalmente ou entender os detalhes de seus mecanismos. Observa-se ainda que mensurações cuidadosas de efeitos causais freqüentemente conduzem a um melhor entendimento do mecanismo causal envolvido (Holland & Rubin, 1988). Mais especificamente, o objetivo da estatística na questão causal tem-se concentrado no estudo de efeitos causais relativos a possíveis manipulações que são previamente estabelecidas. Essa abordagem segue um caminho contrário a uma substancial discussão não estatística de causa e efeito, que se concentra em estimar (ou mesmo determinar) qual a causa de uma particular resposta (Rubin, 1990a). Essa idéia deverá ficar mais clara ao longo dos capítulos deste livro.

Inferência causal em epidemiologia, apesar de conter o termo 'inferência', bastante peculiar à estatística, não se tem valido, tradicionalmente, de princípios estatísticos para sua avaliação. A tônica da discussão tem-se dividido entre a abordagem filosófica e o estabelecimento de condições ou restrições que respaldem uma interpretação causal. Nesse aspecto, a principal contribuição da estatística refere-se a uma dessas condições, que é a verificação da existência de associação estatística entre a variável resposta (doença) e o suposto fator causal (exposição). Desse modo, apesar de a teoria estatística nos últimos anos ter subsidiado um substancial desenvolvimento da metodologia epidemiológica por meio de técnicas quantitativas avançadas, o problema fundamental de inferir causalidade ainda persiste, sobretudo quando investigada com base em estudos observacionais (Breslow, 1996). Mesmo sob randomização,[1] analistas mais cautelosos

[1] Randomização (ou aleatorização) é um mecanismo aleatório de alocação dos tratamentos (causas) às unidades, presente apenas em ambientes experimentais, reconhecido como apropriado para obtenção de conclusões causais.

procuram falar sobre causalidade com ressalvas. Nosso objetivo, aqui, é rever um modelo estatístico de causalidade, abordando tópicos e contribuições recentes ou ainda não efetivamente incorporadas ao pensamento e à prática epidemiológicos.

Para a tarefa de identificação de fatores causais, a epidemiologia tem-se utilizado tanto de estudos observacionais quanto de experimentais (os ensaios clínicos). Uma vez que a experimentação nem sempre é possível quando se trabalha com populações humanas, maior ênfase tem sido dada aos estudos observacionais, principalmente aos estudos de caso-controle cuja execução, embora mais sujeita a vieses, torna-se, eventualmente, a única opção viável em razão de questões operacionais tais como tempo, custo, raridade de uma determinada doença etc. Entretanto, a possibilidade de se estabelecer um modelo estatístico cuja formulação seja comum às abordagens experimental e observacional parece adequada, pois unificaria a idéia de causa contida nos estudos epidemiológicos. A diferença fundamental entre as abordagens se concentra no nível de controle que o analista, experimentador ou observador possui sobre o mecanismo de determinação de qual tratamento (causa) a unidade recebe.

Independentemente de qual desenho epidemiológico, experimental ou observacional um pesquisador utilize, uma proposição causal a ser investigada – tal como 'tabagismo causa doença cardiovascular' – apresenta características que dificultam sua avaliação. Não apenas é evidente que nem todos os fumantes apresentarão doença cardiovascular, como também é evidente que alguns não-fumantes poderão desenvolver a doença. A proposição tem sido avaliada, portanto, incorporando a idéia de risco. Se, de fato, fumar tem algum efeito, seria esperado encontrar um risco maior de ocorrência de doença cardiovascular entre fumantes do que entre não-fumantes. Nessa perspectiva, o papel do mecanismo que designa (determina) o 'tratamento' (no caso, exposição ao fumo) é fundamental. Saber como os indiví-

duos foram alocados nos grupos fumante e não-fumante é de capital importância para que o resultado da comparação entre as taxas (ou riscos) encontradas possa ser atribuído à causa em questão, e não a uma explicação alternativa. Por exemplo, se os fumantes são mais idosos que os não-fumantes, a que se deveria atribuir uma eventual elevação na taxa de doença cardiovascular quando se comparassem fumantes e não-fumantes? Ao hábito de fumar ou à idade? Além disso, os métodos de inferência estatística usam o mecanismo de designação para deduzir inferência causal e, para determinados dados e suposições, as inferências causais eventualmente variam quando este mecanismo varia (Rubin, 1991).

Apesar da razoável idade das primeiras tentativas metodológicas para identificação das causas das doenças, a atualidade do tema ainda é marcante. Os clássicos trabalhos de Rothman (1976) e Susser (1977) foram merecedores de recentes reedições (Wilkelstein, 1995; Greenland, 1995). A abordagem aqui apresentada é fortemente sustentada no já clássico trabalho de Holland (1986), que cunhou o termo 'Modelo de Rubin', explorando a noção de causa por meio de uma lógica contrafactual. Uma revisão desse modelo aparece em Little & Rubin (2000). Entretanto, só mais recentemente surgiram algumas aplicações em epidemiologia. Os trabalhos de Rubin (1991), Efrom & Feldman (1991) e Halloran & Struchiner (1995) são referências da utilização, em problemas epidemiológicos, da noção de causalidade devida a Rubin. Embora a visão de causalidade sob uma ótica contrafactual tenha se desenvolvido e sido defendida (Greenland, 2000), ainda não há um consenso por parte dos autores (Dawid, 2000).

A proposta é, portanto, apresentar a contribuição da estatística para a discussão sobre causalidade por meio da incorporação das idéias de Rubin e outros autores, explorando sua aplicabilidade aos principais desenhos epidemiológicos e ressaltando a importância do mecanismo de designação das causas consideradas. Além disso, mais que um objetivo, tem-se a pre-

tensão de estreitar ainda mais as relações entre o estatístico e o epidemiologista. Se, de um lado, um estatístico tradicionalmente se esquiva quando confrontado com uma questão causal, de outro, um epidemiologista não tem contato com uma noção estatística de causa. Construir com base nesse conceito uma ponte entre estas duas áreas, estatística e epidemiologia, parece ser bastante interessante para se ressaltar ainda mais a subárea conhecida como bioestatística.

Os capítulos e seções aqui apresentados não são estanques. Randomização, confundimento  e permutabilidade e o conceito de validade, por exemplo, apesar de constituírem tópicos separados, se relacionam estreitamente quando o tema em questão se refere à noção de causa. Assim, quando se abordar determinado tópico, alguns outros estarão eventualmente presentes, direta ou indiretamente.

# 1

# Causalidade e Epidemiologia

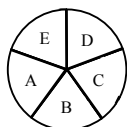
A epidemiologia, como ciência preocupada com a freqüência, a distribuição e os 'determinantes' das doenças que acometem a população, tem desenvolvido procedimentos metodológicos baseados em modelos estatísticos que buscam identificar a etiologia das doenças. Esses modelos são, entretanto, dependentes de pressupostos que muitas vezes não podem ser checados com base em dados observados. O conceito de validade tem, portanto, um papel-chave na avaliação dos efeitos causais. Por sua vez, a validade sobre a existência de uma relação de causa e efeito entre uma doença e um fator de risco é dependente das características de cada desenho de estudo que a epidemiologia utiliza.

Segundo Rothman & Greenland (1998), uma causa pode ser entendida como qualquer evento, condição ou característica que desempenhe uma função essencial na ocorrência da doença. Observa-se, ainda, que causalidade é um conceito relativo, devendo ser compreendido em relação a alternativas concebíveis. Isto é, o efeito de uma causa é sempre relativo a uma outra causa. A expressão 'A causa B' significa que A é a causa de B relativa a alguma outra causa que, freqüentemente, se refere à condição 'não A' (Holland, 1986). Por exemplo, ao se falar que história de tabagismo inveterado é uma causa para câncer de pulmão, é necessário especificar a causa alternativa, que pode ser, por exemplo, tabagismo recente ou não tabagismo.

Ainda em relação à questão conceitual de causa, a epidemiologia tem trilhado um caminho que passa pela filosofia da ciência. São vários os auto-

res que seguem essa linha. O trabalho de Rothman (1988) agrega ensaios e comentários de diversos autores sobre causalidade sob a ótica filosófica, discutindo questões ligadas à lógica da causalidade e à teoria da refutação de Popper. Entretanto, o processo de identificação de um suposto agente causal pode ser simplificado utilizando-se uma abordagem mais pragmática, sem a necessidade de um aprofundamento na questão teórico-conceitual intrínseca à noção de causa. Além disso, enquanto os cientistas em geral consideram associações causais como etapas do processo de conhecimento da epidemiologia e da história natural da doença, profissionais envolvidos diretamente com a prevenção das doenças necessitam, para suas ações, de conclusões rápidas, tão logo alguma evidência tenha sido atingida.

A questão causal no ambiente epidemiológico tem sido apresentada também de uma forma determinística, observando-se, entretanto, que a ocorrência de uma doença em geral não está associada exclusivamente a uma única causa. Para ocorrência da doença, é necessário um conjunto de causas componentes. Rothman & Greenland (1998) definem causa suficiente como um conjunto de eventos e condições mínimos que inevitavelmente acarreta a ocorrência de doença, no qual 'mínimo' implica que não se pode prescindir de nenhum dos eventos ou condições componentes. Nota-se ainda que, para a ocorrência de uma determinada doença, pode haver diversos conjuntos de causas suficientes. Algumas causas componentes, quando presentes em todas as causas suficientes alternativas, são chamadas causas necessárias. Outras, para serem identificadas, dependem da interação com outras causas componentes. A Figura 1 ilustra esse modelo, no qual se observam três causas suficientes, sendo A uma causa necessária. Um modelo de causalidade com essas características, isto é, fundamentado na classificação dos mecanismos que precedem a resposta, diferentemente do modelo a ser discutido, que se baseia na classificação de respostas individuais à exposição, tem sido conhecido como modelo de Rothman ou modelo de causas suficiente/componente (Greenland, 1995).

Figura 1 – Exemplo do modelo de causalidade de Rothman para uma particular doença

**Causa Suficiente I** **Causa Suficiente II** **Causa Suficiente III**

Muitas causas que são de interesse da epidemiologia, embora não sejam suficientes, são componentes de causas suficientes. Dispor de água não tratada não é suficiente para o surgimento de doenças diarréicas e fumar não é suficiente para produzir câncer de pulmão, mas ambas são causas componentes de causas suficientes. Observa-se, ainda, que a identificação completa de todas as causas componentes de uma determinada causa suficiente, mesmo que seja possível e viável, não é fundamental se o objetivo é a prevenção da doença. Por exemplo, mesmo não sendo capaz de identificar todas as causas componentes de uma dada causa suficiente para câncer de pulmão, entre as quais está o hábito de fumar, é possível prevenir aqueles casos que resultariam desta causa suficiente pela remoção do fumo da constelação de causas componentes (Rothman, 1976).

O fato é que a maioria ou mesmo todos os componentes de uma causa suficiente são desconhecidos. Exigem-se, portanto, hipóteses específicas e modelos apropriados para que a avaliação dos efeitos observados possa ser atribuída a uma causa estabelecida. Além disso, o conhecimento biológico sobre uma proposição epidemiológica é freqüentemente raro, tornando-a, na maioria das vezes, uma mera afirmação de associação entre a exposição e a doença. Averiguar a existência de associação é tarefa bem familiar à estatística. A passagem da atribuição de associação entre a exposição (suposto agente causal) e a doença para atribuição de causa tem sido a tônica da discussão

sobre causalidade. A partir da incorporação das idéias de Rubin, traduzidas no chamado Modelo de Rubin, uma nova lógica pode ser desenvolvida.

Historicamente, a primeira tentativa formal para identificação das causas de uma doença se deu com a formulação, em 1890, do que foi chamado de Postulados de Henle-Koch (Evans, 1978). Tais postulados satisfaziam a necessidade de se estabelecer regras que guiassem a investigação de bactérias como possíveis agentes causais (Yerushalmy & Palmer, 1959). Basicamente, estabeleciam os seguintes critérios para o organismo patogênico:

- deve estar presente em todos os casos da doença em questão;
- não deve ocorrer nem de forma casual nem de forma patogênica em outra doença;
- isolado do corpo e crescido em cultura pura, deve induzir a doença quando inoculado em suscetíveis.

Foi observado, inclusive por Koch, que para a identificação de um suposto agente causal não era necessário que todos os critérios fossem satisfeitos e que apenas os dois primeiros já eram suficientes. Ou seja, o cumprimento dos postulados fornecia razoáveis elementos para aceitar a hipótese de causalidade do suposto agente, enquanto o seu descumprimento não deveria excluir tal hipótese.

Apesar de suas limitações, que os tornavam não aplicáveis a todas as doenças bacterianas e eventualmente não aplicáveis às doenças viróticas e parasitárias, esses postulados serviram de base para que a discussão sobre a etiologia das doenças, considerando-se suas especificidades, fosse ampliada. Assim, à medida que o conhecimento sobre as doenças crescia, como, por exemplo, identificando-se novos vírus e suas respectivas características biológicas, tempo de incubação e possibilidade de imunidade, os postulados de Henle-Koch foram sendo testados e modificados. As modificações consistiam, basicamente, na incorporação de novas restrições para que a atribuição de causalidade servisse a todas as doenças, inclusive as crônicas.

As modificações culminaram com os critérios estabelecidos por Hill em 1965. Identificada uma associação entre exposição e doença, ele sugeriu que os seguintes aspectos fossem considerados na tentativa de se distinguir uma associação causal de uma não-causal:

- Força da Associação – uma associação será tão mais forte quanto mais distante do valor de nulidade estiver a medida de efeito de interesse calculada.[2] O argumento é que uma associação forte tem mais chance de ser causal do que uma associação fraca. Isso porque se ela se deveu a algum viés; então, muito provavelmente, este viés seria evidente. Associações fracas, contudo, são mais prováveis de serem explicadas por vieses não detectados. Entretanto, uma associação fraca não descarta a hipótese de causalidade. Outra característica importante é que a força de uma associação não é um aspecto compatível biologicamente, e sim uma característica que depende da prevalência das outras causas componentes. Uma forte associação de determinada causa pode indicar simplesmente que esta causa é pouco prevalente em relação às outras e não possuir, assim, significado biológico importante (Rothman & Greenland, 1998).

- Consistência – a consistência se refere à repetição dos achados para diferentes populações. Resultados similares reforçam a hipótese de causalidade.

- Especificidade – uma causa é específica para um determinado efeito se a introdução de um suposto fator causal é seguido da ocorrência do efeito e sua remoção implica que tal efeito não ocorra. Em razão do fato de que muitos fatores implicam muitos efeitos e praticamente todas as doenças têm múltiplas causas, a especificidade de uma asso-

[2] Uma medida de efeito é uma comparação (diferença ou razão) entre medidas de freqüência de doença (prevalências, riscos, taxas ou *odds*) calculada para dois grupos, freqüentemente expostos e não expostos a determinado fator sob investigação causal.

ciação respalda uma interpretação causal, mas sua falta não deve ser indicação de não-causalidade.

• Temporalidade – a causa deve necessariamente preceder o efeito. Os estudos transversais e retrospectivos muitas vezes carecem dessa evidência, dificultando uma atribuição de causalidade.

• Gradiente Biológico – este aspecto refere-se à presença de uma curva dose-resposta. Observar uma freqüência crescente de ocorrência de doença à medida que se aumenta a dose ou o nível da exposição reforça a hipótese de causalidade. Entretanto, a observação de um efeito dose-resposta pode ser devida completamente a algum viés.

• Plausibilidade – se o efeito hipotetizado é plausível diante do conhecimento biológico vigente, a interpretação causal é fortalecida. No entanto, plausibilidade biológica não pode ser exigida, uma vez que depende do conhecimento disponível à época da investigação. Em geral, quanto menos se conhece a respeito da etiologia da doença e doenças similares, menos segurança se tem para rejeitar uma interpretação causal com base neste critério (Kleibaum et al., 1982).

• Coerência – o critério de coerência é satisfeito quando a associação encontrada não entra em conflito com o que é conhecido sobre a história natural e a biologia da doença. Nota-se que este critério combina aspectos dos critérios de consistência e plausibilidade biológica. Ele mereceu especial atenção em recente trabalho (Rosembaun, 1994), no qual se buscou quantificar a evidência fornecida por uma associação coerente.

• Evidência Experimental – é conhecido o poder da experimentação na avaliação de causalidade. Entretanto, a obtenção de tal evidência é raramente disponível em estudos envolvendo populações humanas devido, principalmente, a questões éticas.

- Analogia – uma analogia simples pode aumentar a credibilidade para uma atribuição de causalidade. Por exemplo, se é conhecido que certa droga causa má-formação congênita, talvez uma outra similar que se está estudando também poderia, por analogia, apresentar o mesmo efeito.

À exceção do critério de temporalidade, nenhum outro desses nove critérios de evidência epidemiológica sugeridos por Hill (1965) deve ser exigido como condição *sine qua non* para julgar se uma associação é causal. Pode-se dizer também que eventualmente os critérios de evidência experimental e analogia são irrelevantes e o de especificidade, impróprio. Holland (1986) agrupa os critérios de plausibilidade, coerência e analogia por entender que os três se referem aos conhecimentos adquiridos até a época do estudo. Essa abordagem é, pois, condicionada a aspectos e critérios que na maioria das vezes não são nem necessários nem suficientes. Quando são satisfeitos reforçam a hipótese de causalidade, mas quando isso não ocorre, não se deve descartá-la.

De modo diferente, a abordagem estatística sobre causalidade baseia-se na formulação de um modelo construído sob uma ótica contrafactual,[3] na qual são estabelecidas hipóteses que procuram viabilizar a inferência causal. A dificuldade está, pois, na verificação de tais hipóteses, nem sempre passíveis de serem testadas por meio dos dados observados. A validade de algumas hipóteses não testáveis, por sua vez, depende do nível de convencimento que o analista consegue obter, para si próprio e para os outros, com base em uma cuidadosa análise de cada situação em particular. Sendo assim, é de grande valia que as hipóteses não testáveis adjacentes ao modelo estejam explicitadas, para que possam ser analisadas criticamente.

[3] Construído sob uma ótica contrafactual significa que o modelo é fundamentado em valores hipotéticos, isto é, em valores que não são efetivamente observados.

## 2

# O Modelo Estatístico de Causalidade

## Causalidade e Estatística

A relação entre causa e estatística merece alguns comentários iniciais. Sustentados na retórica estatística de que correlação não implica causalidade e apesar do reconhecimento geral de que um experimento randomizado bem planejado pode fornecer uma poderosa ajuda na investigação de relações causais, os estatísticos abstiveram-se de pensar em causalidade de forma mais específica e audaciosa. Isso é sutilmente evidenciado quando, por exemplo, lemos o trabalho de Kempthorne (1978), um grande estatístico contemporâneo, que, muito modestamente, se preocupa em não ser arrogante quando ele, um 'mero' estatístico (em sua auto-avaliação), aborda a questão do que seja um conceito viável de causa.

Dempster (1990) argumenta que a primeira razão para que os estatísticos analisem causalidade é prática. O pensamento causal está profundamente presente no entendimento científico dos problemas da estatística aplicada. Por meio de alguns exemplos práticos, tal como a questão de saber se a redução da ingestão de álcool por mulheres elimina ou adia o surgimento de alguns casos de câncer de mama, o autor conclui que a coleta e a análise de dados estatísticos constituem uma porção apreciável da base empírica em que ocorrem as tentativas cientificamente confiáveis de atribuição de causalidade. Ou, como observado por Gail (1996), o pensamento estatístico,

presente na coleta e na análise de dados, tem sido crucial para o entendimento do poder e das potenciais fragilidades das evidências científicas. O resultado disso é um crescente desenvolvimento do horizonte do estatístico e dos métodos estatísticos, impulsionando o surgimento de novas metodologias e debates construtivos sobre critérios necessários para se inferirem relações causais. Esse conhecimento estatístico forma a base de grande parte da prática epidemiológica atual.

Em termos mais específicos, o processo de inferir causalidade envolve diversas tarefas estatísticas. É decisivamente importante, quando se discutem efeitos causais, começar definindo a quantidade a ser estimada. Subseqüentemente, deve-se estudar métodos para coletar dados com os quais essa quantidade possa ser estimada. E, finalmente, deve-se considerar técnicas estatísticas para inferir os efeitos com base nos dados observados (Rosenbaun & Rubin, 1984). É na etapa de definição da quantidade a ser estimada que se concentra a mudança na lógica da investigação causal devida a Rubin (1974), por meio do modelo de respostas potenciais. Entretanto, no momento da coleta de dados para estimação dessa quantidade, o esquema de amostragem necessário deve ser considerado em razão do desenho de estudo epidemiológico. Já para a etapa de inferência causal baseada nas observações, diversas técnicas podem ser utilizadas. Rubin (1990a) tem discutido e comparado essas técnicas, com especial destaque para o método bayesiano.

Assim, o problema de atribuição de causalidade, em uma leitura estatística, é interpretado como um problema de detecção de efeitos causais por meio de mensurações. Medir efeitos causais constitui, pois, a base do desenvolvimento da teoria estatística que discute causalidade, com particular interesse em epidemiologia. Nessa visão, inferir estatisticamente efeitos causais traduz-se em um processo de estimação de efeitos devido a potenciais manipulações que podem ou poderiam ser aplicadas a uma unidade em um con-

texto real com todas as suas complexidades, necessitando, portanto, de definições adequadas a essa estrutura complicada (Rubin, 1990a).

Uma primeira característica importante associada à contribuição da estatística para a discussão de causalidade, que acaba se refletindo em uma limitação, se refere à concentração do estudo nos efeitos de causas previamente determinadas, em oposição à abordagem usual, que busca as causas de efeitos observados. Essa é uma condição imposta por Holland (1986) e defendida por Rubin, o qual, de forma incisiva, expressa:

> Em qualquer situação relativamente complicada do mundo real, envolvendo plantas, animais, pessoas, aviões ou reatores nucleares, eu acredito que é geralmente impossível examinar uma resposta observada e realisticamente encontrar a causa dela. A única esperança para tal atribuição é descrever a situação muito cuidadosamente para limitar o tratamento (causa) sugerido que tenha ocorrido e indicar a) quais outros fatores (outras potenciais causas) estão sendo assumidos fixados em seus valores observados; e b) quais tratamentos (causas) alternativos contrafatuais estão sendo considerados terem acontecidos preferivelmente ao tratamento observado, que é a causa postulada. (Rubin, 1990a:280)

Dempster (1990) tenta justificar essa insistência de Holland (1986) e Rubin (1990a), afirmando que os princípios estatísticos que estudam as dificuldades de se fazer induções sobre o mundo real com base em correlações empíricas não estão estreitamente ligados ao conceito de causa. Acrescenta que o estatístico deve também estar interessado na abordagem tradicional, que estuda as causas de efeitos observados, utilizando-se, entretanto, de princípios complementares àqueles descritos por Holland no Modelo de Rubin. Entretanto, essa restrição parece justificável, uma vez que tenta responder de forma mais pragmática à questão de como identificar uma causa. Não necessariamente exclui outras formas de estudar causalidade, apenas limita a participação da estatística nessa discussão. Os princípios complementares aludidos por Dempster (1990) fazem uso mais especificamente de conceitos próprios à teoria filosófica.

Outra característica importante contida nesta discussão, e que também é bastante controversa, se refere à definição do que pode ser considerado uma causa. A condição-chave dessa idéia, sob a ótica estatística, é que cada unidade deve estar potencialmente exposta à ação de qualquer uma das causas cujo efeito poderia ser medido. Assim, atributos pessoais imutáveis, tais como sexo ou raça, não podem ser vistos como causas, já que não se poderia observar seu efeito sob a condição alternativa àquela que a unidade possui. É bastante comum encontrar declarações de causalidade atribuídas a características pessoais. Mas tais declarações, na conotação de causa de Holland (1986), são sempre declarações de associação entre os valores de um atributo e uma variável resposta para as unidades de uma população. Assim, pode-se dizer, por exemplo, que certa pessoa não teve câncer de pele (melanoma) 'porque' era negra. No entanto, tudo o que pode ser inferido dessa declaração é que a proporção de casos de melanoma é inferior entre negros, apesar da condição causal que a declaração carrega. Tal restrição está longe de ser um consenso por parte de autores preocupados com a questão da causalidade, principalmente os filósofos. Entretanto, entre os estatísticos ela é razoavelmente bem aceita. Kempthorne (1978) diz que é sem sentido epistemológico falar de uma característica de um indivíduo 'causando' ou determinando uma outra característica individual. Cox (1986, 1992) parece acolher a idéia ao reconhecer que certas variáveis não podem propriamente ser vistas como causas.

Deve ser reconhecido, todavia, que se um atributo pessoal pode ou não ser uma causa depende da conotação que se dá à palavra. Mas, para o modelo estatístico a ser discutido a seguir, essa condição é teoricamente importante, pois a idéia de causa nele contida se refere à comparação de respostas potenciais sob as diferentes situações de exposição (causas ou tratamentos) consideradas.

# O Modelo de Respostas Potenciais

## Estrutura do modelo

Um estudo estatístico de efeitos causais é, portanto, aquele em que se comparam os resultados de dois ou mais tratamentos em uma população de unidades, onde, em princípio, cada uma destas unidades poderia ser exposta a qualquer um dos tratamentos. A estruturação desse modelo, também conhecido como Modelo de Rubin, foi pioneiramente descrita de forma eloqüente por Holland (1986) e apresentada novamente de forma sucinta em outros trabalhos (Holland & Rubin, 1988; Holland, 1989; Wainer, 1991; Halloran & Struchiner, 1995; Angrist et al., 1996; Little & Rubin, 2000). A seguir, tenta-se reproduzir quase que integralmente essa estrutura, a fim de consolidar a notação.

Os elementos essenciais que compõem o Modelo de Rubin:

- uma população de unidades, $U$;
- um conjunto, $K$, de agentes causais bem definidos (também chamados tratamentos ou causas) para os quais cada unidade $u \in U$ possa ser exposta. Para efeito de simplificação, serão considerados apenas dois agentes causais, $K = \{t; c\}$, onde $t$ representa tratamento ou exposição e $c$ controle ou não exposição;
- uma resposta, $Y$, a variável dependente, que pode ser registrada para cada unidade após a exposição aos agentes causais em $K$. $Y$ será considerada dicotômica, na medida em que em epidemiologia o principal interesse é a ocorrência ou não de uma determinada doença. A extensão para um $Y$ qualquer é imediata.

Nesse modelo, o papel do tempo é fundamental. Quando uma unidade é exposta a uma causa, isso deve acontecer em algum tempo ou dentro de um período de tempo específico. Assim, as variáveis dividem-se em pré-

exposição, aquelas cujos valores são determinados anteriormente à exposição à causa, e pós-exposição, aquelas determinadas após. A função da variável resposta $Y$ é medir o efeito de uma causa. Logo, se encontra na classe das variáveis pós-exposição. Isso dá surgimento a uma característica crítica do modelo, isto é, o valor de uma variável pós-exposição é potencialmente afetado pela particular causa, $t$ ou $c$, para a qual a unidade é exposta. E isso é exatamente equivalente à declaração de que causas têm efeitos, que é a essência da idéia de causalidade.

Assim, no lugar de uma variável dependente simples $Y$, tem-se uma variável dependente $Y_k$ para cada um dos tratamentos para os quais a unidade pode potencialmente ser exposta. Se uma unidade é exposta ao agente causal $t$, registra-se o valor $Y_t$ para esta unidade. Se a mesma unidade tivesse sido exposta ao agente causal $c$ no lugar de $t$, seria registrado o valor $Y_c$ e não o valor $Y_t$. Formalmente, para dois tratamentos, associa-se o vetor $(Y_t, Y_c)$ para cada unidade $u \in U$, onde $Y_k(u)$ é a resposta obtida para a unidade $u$ quando exposta à causa $k \in K$.

O efeito da causa $t$ sobre $u$ quando medido por $Y$ relativo à causa $c$ é a diferença entre $Y_t(u)$ e $Y_c(u)$. No modelo, será representado pela diferença algébrica

$$Y_t(u) - Y_c(u).$$

Holland (1986) chama essa diferença de efeito causal de $t$ relativo à $c$ sobre $u$, quando medido por $Y$. Essa é a maneira com que o modelo para inferência causal expressa a mais básica de todas as declarações causais, sendo na realidade a quantidade que se gostaria de poder estimar a partir de dados observáveis e hipóteses plausíveis. Uma forma alternativa de se quantificar o efeito causal é considerar a razão entre as respostas potenciais.

Assim, dentro da situação considerada, onde tanto a variável exposição (tratamento) quanto a variável resposta (doença) são dicotômicas, assumindo-se o valor 1 para presença e 0 para ausência, o efeito causal indi-

vidual $Y_t(u) - Y_c(u)$ será igual a 1 somente se *u* torna-se doente quando exposto, mas não se torna doente quando não exposto. $Y_t(u) - Y_c(u)$ será 0 quando o *status* de doença de *u* é o mesmo, independentemente da condição de exposição. E, finalmente, será -1 se *u* não adoece quando exposto, mas adoece quando não exposto. Assim, para um determinado indivíduo $u_0$, o valor 0 para o efeito causal $Y_t(u) - Y_c(u)$ é indicação de ausência de efeito da causa *t* relativa à causa *c* para a unidade $u_0$. Já os valores 1 e -1 representam existência de efeito da causa postulada, sendo o valor -1 indicativo de uma causa preventiva – por exemplo, uma vacina.

Inferência causal está, enfim, interessada no efeito de causas sobre unidades específicas, isto é, está interessada em determinar o valor do efeito causal $Y_t(u) - Y_c(u)$. É frustrada, entretanto, por uma limitação de observação que Holland (1986) chama de Problema Fundamental da Inferência Causal. Qual seja: é impossível 'observar' os valores de $Y_t(u)$ e $Y_c(u)$ para a mesma unidade *u* e, portanto, é impossível 'observar' o efeito de *t* relativo a *c* sobre *u*. No entanto, a aplicação dessa afirmação depende da natureza dos tratamentos e das unidades sob estudo.

A aparente inviabilidade de fazer inferências causais em conseqüência desse problema desaparece quando se nota que a impossibilidade de observação simultânea de $Y_t(u)$ e $Y_c(u)$ não significa ausência total de informação relevante sobre estes valores. E essa informação depende da situação considerada. Holland (1986) assinala duas soluções para o problema, que chama de 'solução científica' e 'solução estatística'.

A solução científica faz uso de hipóteses de homogeneidade ou estabilidade. Ao estudar o comportamento de uma peça em laboratório, um cientista pode acreditar que o valor de $Y_c(u)$ não depende do momento em que é medido – hipótese de 'estabilidade temporal' – e o valor de $Y_t(u)$ não é afetado pela exposição anterior da unidade *u* à causa *c* – hipótese de causa 'transiente'. Assim, para superar o Problema Fundamental da Inferência

Causal, o cientista expõe $u$ à causa $c$, mede $Y_c(u)$ e, subseqüentemente, expõe $u$ à causa $t$ e mede $Y_t(u)$. A obtenção do efeito causal em nível individual $Y_t(u) - Y_c(u)$ torna-se, então, imediata. Note-se, entretanto, que o efeito causal que pode ser obtido pelo cientista é dependente de hipóteses não testáveis. Isto é, o cientista não é capaz de provar as suposições que fez de estabilidade temporal e causa transiente. Com um trabalho cuidadoso, ele pode convencer a si e aos outros de que são válidas, mas nunca poderá estar absolutamente certo disso. Essa abordagem é bastante comum em experimentos físicos e está presente em nossas pequenas avaliações de causalidade feitas diariamente.

Uma segunda maneira de aplicar a solução científica é assumir que $Y_t(u_1) = Y_t(u_2)$ e $Y_c(u_1) = Y_c(u_2)$ para duas unidades $u_1$ e $u_2$. Ou, de forma mais geral, assumir que $Y_k(u_1) = Y_k(u_2)$ para todo par de unidades $u_1$ e $u_2$. Essa é a hipótese que Holland (1986) chama de 'homogeneidade de unidades'. Sob tal hipótese, a obtenção do efeito causal em nível individual também se torna imediata, uma vez que $Y_t(u) - Y_c(u) = Y_t(u_1) - Y_c(u_2)$ para todo $u$, $u_1$ e $u_2 \in U$. Apesar de também ser bastante comum em laboratórios, essa solução para o problema fundamental da inferência causal tem sido buscada por estatísticos e epidemiologistas, que procuram operacionalizá-la por meio de procedimentos de seleção de unidades. A maneira que os cientistas de laboratórios (e, de forma análoga, os epidemiologistas) utilizam para se convencer de que as unidades são homogêneas é prepará-las (ou selecioná-las) cuidadosamente, de modo que elas pareçam idênticas em todos os aspectos relevantes. Apesar de se poder fazer a hipótese de homogeneidade plausível, sua validade não é passível de ser testada principalmente devido à possibilidade de existência de outras variáveis não observadas eventualmente importantes. Na maioria dos estudos epidemiológicos, senão em todos, a hipótese de homogeneidade obtida por estratificação, isto é, pela separação das unidades em subgrupos com base em co-variáveis observadas, dificilmente seria justificada, pois raramente

existiriam situações em que todos os fatores de risco importantes são acuradamente medidos e controlados (Greenland, 1990). Entretanto, a estratificação e, num caso mais extremo, o pareamento constituem estratégias de seleção de unidades comumente utilizadas na avaliação de efeitos causais por meio de uma aproximação, tanto quanto possível, da hipótese de homogeneidade de unidades.

Como visto, estratificação é entendida como a busca de unidades homogêneas por meio da identificação de subpopulações com base em co-variáveis pré-exposição observadas. O pareamento é uma tentativa de se atingir comparabilidade entre potenciais fatores de confundimento no estágio de desenho do estudo. Isso é feito selecionando-se apropriadamente, para o estudo, unidades para formar os pares que sejam tão semelhantes quanto possível com respeito a potenciais variáveis de confundimento. Consiste, portanto, na escolha de uma ou mais unidades controle para cada unidade tratada que seja similar quanto a características mensuradas anteriormente à ação dos tratamentos.

Cabe aqui uma pequena apresentação da idéia de confundimento. Uma discussão mais detalhada será feita em tópico específico. Como exemplo, considere um estudo experimental para avaliação de efeitos causais no qual, uma vez analisados os dados apropriadamente a partir da amostra disponível, conclui-se pela existência de associação entre exposição e doença, não importando por ora a magnitude desta associação. Três situações alternativas devem ser consideradas:

- existe de fato uma associação causal entre exposição e doença e o estudo foi capaz de detectá-la;
- 'não' existe de fato associação causal entre exposição e doença, mas por mero acaso uma 'infeliz' amostra forneceu evidências de causalidade. Neste caso, a ordem de grandeza do acaso é comumente medida pelo p-valor;

- 'não' existe de fato associação causal, mas um viés no estudo foi responsável pela evidência amostral de causalidade quando avaliada pelo p-valor.

Em epidemiologia, os principais vieses, cujo controle é fundamental para causalidade, têm sido tradicionalmente classificados em viés de seleção, viés de informação e confundimento (ou confusão). Assim, confundimento se refere a uma explicação alternativa para uma determinada conclusão, constituindo portanto um dos principais – senão o principal – problemas para a inferência causal. Aproveitando o exemplo para antecipar um outro elemento fundamental desse modelo estatístico de causalidade, esta terceira situação, em que os dados apontam a existência de causalidade em virtude da presença de algum viés, tem sido apresentada também como um problema relacionado ao mecanismo de designação de tratamentos (Rubin, 1991). Assim, confundimento poderia ser visto também como algo devido ao emprego de um mecanismo de designação de tratamentos impróprio.

Voltando ao Problema Fundamental da Inferência Causal, de forma diferente da solução científica, a solução estatística substitui a busca do efeito causal obtido em nível individual pela busca de um efeito causal 'típico', em nível populacional ou subpopulacional. O efeito causal médio, *T*, de *t* relativo à *c* sobre *U* é a esperança da diferença $Y_t(u) - Y_c(u)$ para as unidades $u \in U$. Isto é,

$$T = \mathrm{E}(Y_t - Y_c).$$

Que, pelas regras usuais de probabilidade, pode ser expresso como

$$T = \mathrm{E}(Y_t) - \mathrm{E}(Y_c).$$

Essa expressão revela que respostas que 'podem ser observadas' sobre diferentes unidades podem ser utilizadas para obter informação sobre *T*. Se algumas unidades são expostas à causa *t*, elas podem ser usadas para fornecer informação sobre $\mathrm{E}(Y_t)$, já que este é o valor médio de $Y_t$ sobre *U*, enquanto outras unidades podem ser expostas à *c* e se obtiver informação sobre $\mathrm{E}(Y_c)$.

Na situação considerada, em que a resposta $Y$ é dicotômica assumindo o valor 1 quando presente e 0 quando ausente, pode-se reconhecer no parâmetro $T$, definido anteriormente, a medida de efeito conhecida pelos epidemiologistas como risco atribuível. Isso porque

$$\mathrm{E}(Y_t) = 1.\Pr(Y_t = 1) + 0.\Pr(Y_t = 0) = \Pr(Y_t = 1) = \text{Risco entre tratados} = R_t$$

De forma análoga, o risco entre controles (ou não expostos) é

$$R_c = \Pr(Y_c = 1) = \mathrm{E}(Y_c).$$

Assim, $T$ mede a parcela do risco entre os tratados que pode ser atribuída ao tratamento, já que remove o risco em conseqüência de outras causas.

O mecanismo que indica qual tratamento (exposição), $t$ ou $c$, a unidade $u$ recebe, é muito importante e envolve todas as considerações estatísticas de um experimento bem planejado, tal como a randomização ou procedimentos a ela alternativos. Assim, a solução estatística substitui o efeito causal impossível de se observar de $t$ sobre unidades específicas, pelo possível de ser estimado efeito causal médio de $t$, sobre uma população de unidades. Em resumo, diante da impossibilidade de se observar $Y_t(u)$ e $Y_c(u)$ para a mesma unidade, elege-se um conjunto de unidades a serem expostas à causa $t$ estimando-se $\mathrm{E}(Y_t)$ e um outro conjunto a ser exposto à causa $c$ para se estimar $\mathrm{E}(Y_c)$.

Tal como a solução científica, a solução estatística para o Problema Fundamental da Inferência Causal pressupõe a validade de hipóteses não passíveis de serem testadas, ou apenas parcialmente testáveis. As hipóteses adjacentes à solução estatística são duas: a de independência ou a de efeito constante. Para discussão da hipótese de independência, é necessário incorporar um novo elemento ao modelo. Seja $S$ a variável que indica a causa (tratamento) a que cada unidade em $U$ é exposta. Isto é, $S(u) = t$ indica que a unidade é exposta à causa $t$ e se observa $Y_t(u)$. Por sua vez, $S(u) = c$ indica exposição à $c$ e a resposta observada é $Y_c(u)$. Assim, a resposta observada para a unidade $u$ é $Y_{S(u)}(u)$. A variável resposta observada é, portanto,

$Y_s$ e, mesmo que agora o modelo contenha três variáveis, $S$, $Y_t$ e $Y_c$, o processo de observação envolve apenas duas, $S$ e $Y_S$. Holland (1986) observa ainda que a distinção entre as três situações envolvidas, isto é, o processo de mensuração $Y$ que produz a variável resposta, as duas versões da variável resposta $Y_t$ e $Y_c$ que correspondem a qual causa a unidade é exposta e a variável resposta observada $Y_S$, é muito importante e freqüentemente não está presente em discussões sobre causa. Essa distinção não surge em estudos de simples associação, mas é crucial para análise de causalidade.

Em um estudo experimental, $S$ é construída pelo experimentador, enquanto em um estudo não controlado (observacional) é determinada por fatores além do controle do analista. Em qualquer caso, a característica crítica da noção de causa nesse modelo é que o valor de $S(u)$ para cada unidade 'poderia ter sido diferente'. Estar sob o controle do experimentador, nesse contexto, significa que ele conhece a estrutura probabilística que indicou o tratamento $k \in K$ à unidade $u \in U$, tal como num processo de randomização em que as unidades são alocadas nos grupos de exposição com base em um mecanismo aleatório, independentemente de outras variáveis.

Em estudos observacionais, o analista procura resgatar as vantagens do processo de randomização, no qual se busca comparabilidade entre os tratamentos considerados, por meio da observação de co-variáveis relevantes (associadas ao tratamento ou à resposta) que podem ser controladas no momento do desenho do estudo ou, posteriormente, na análise. Busca-se então substituir a hipótese de independência por hipóteses menos rigorosas de independência 'condicional', onde o condicionamento é determinado por co-variáveis pré-exposição observadas. Assim, incorpora-se ao modelo estatístico geral para inferência causal um conjunto de $q$ co-variáveis observáveis não afetadas pela exposição, isto é, pré-exposição.

A Tabela 1 sintetiza as informações observadas relevantes para se fazer inferência causal em um contexto geral, comumente disponíveis nos

estudos epidemiológicos. Inferir causalidade em um contexto real e complexo torna-se possível ou não dependendo da capacidade de se obter alguma informação sobre os pontos de interrogação desta tabela, isto é, sobre os valores não observados. O estabelecimento de hipóteses plausíveis, embora nem sempre passíveis de serem validadas pelos dados disponíveis, em conjunto com técnicas estatísticas apropriadas constituem importantes instrumentos para estimação ou conhecimento dos valores que teriam sido observados pelas unidades, tivessem elas recebido tratamento alternativo àquele que de fato receberam.

Tabela 1 – Exemplo de dados observáveis em um estudo de 2 tratamentos ($k$=2) em uma população de $U$ unidades com $q$ co-variáveis e 1 variável resposta (2 respostas potenciais)

| Unidades | Co-variáveis $X$ | | | | $S$ | $Y$ | |
|---|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | ... | $X_q$ | | $Y_t$ | $Y_c$ |
| 1 | $x_{11}$ | $x_{12}$ | ... | $x_{1q}$ | $t$ | $y_t(1)$ | ? |
| 2 | $x_{21}$ | $x_{22}$ | ... | $x_{2q}$ | $t$ | $y_t(2)$ | ? |
| 3 | $x_{31}$ | $x_{32}$ | ... | $x_{3q}$ | $c$ | ? | $y_c(3)$ |
| : | : | : | | : | : | : | : |
| : | : | : | | : | : | : | : |
| $u$ | $x_{u1}$ | $x_{u2}$ | ... | $x_{uq}$ | $c$ | ? | $y_c(u)$ |
| : | : | : | | : | : | : | : |
| : | : | : | | : | : | : | : |
| $U$ | $x_{U1}$ | $x_{U2}$ | ... | $x_{Uq}$ | $t$ | $y_t(U)$ | ? |

Essa estrutura permite visualizar que o processo de inferência de causalidade deve estar condicionado aos dados observados e ao padrão de observação dos dados de fato observados e dos dados perdidos ou não observados (*missing*). Os *missing* da Tabela 1, representados pelos pontos de interrogação, são inerentes à própria formulação do Modelo de Rubin, diferentemente de outros tipos de dados perdidos comumente encontrados na prática

devido a circunstâncias indesejáveis, tais como falta de resposta em um inquérito ou censura. Para fins de causalidade, a consideração de ambos os tipos de *missing* é fundamental.

Para fins de uniformização dos procedimentos estatísticos comumente utilizados para análise, pode ser conveniente a adoção de uma notação comum contida em Rosembaun (1995). As $U$ unidades sob investigação podem ser divididas em $M$ estratos ou subpopulações com base em co-variáveis, isto é, com base em características mensuradas anteriormente à designação dos tratamentos. Assim, existem $U_m$ unidades no estrato $m$, $m = 1, 2,..., M$, de modo que $U = \sum_m U_m$. Faz-se $S_m(u)=1$ se a $u$-ésima unidade do estrato $m$ recebe o tratamento $t$ e $S_m(u)=0$ se recebe o tratamento $c$. Escreve-se ainda $n_m$ para o número de unidades tratadas no estrato $m$, de modo que $n_m = \sum_{u=1}^{U_m} S_m(u)$ e $0 \leq n_m \leq U_m$. E, finalmente, designe-se como $\mathbf{S}$ o vetor coluna $U$-dimensional contendo os valores $S_m(u)$ para todas as unidades, isto é,

$$\mathbf{S} = \begin{bmatrix} S_1(1) \\ S_1(2) \\ \vdots \\ S_1(U_1) \\ S_2(1) \\ \vdots \\ S_M(U_m) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_M \end{bmatrix}, \text{onde} \quad \mathbf{S}_m = \begin{bmatrix} S_m(1) \\ \vdots \\ S_m(U_m) \end{bmatrix}$$

Esta notação compreende diversas situações. Se nenhuma co-variável é utilizada para estratificar a população, então existe um único estrato contendo todas as unidades ($M=1$). Se $U_m = 2$ e $n_m = 1$ para $m = 1, 2,..., M$, então fica caracterizado um estudo pareado com base nas co-variáveis, isto é, cada um dos $M$ pares apresenta uma unidade tratada e uma unidade controle. A situação onde $U_m \geq 2$ e $n_m = 1$ para $m = 1,2,...,M$ é conhecida como pareamento com múltiplos controles, ou seja, existem $M$ conjuntos de unidades, cada qual com uma unidade tratada e uma ou mais unidades controle.

Como já comentado, o efeito causal médio $T$ é a diferença entre os dois valores esperados $\mathrm{E}(Y_t)$ e $\mathrm{E}(Y_c)$. Entretanto, os valores observados $(S, Y_S)$ podem apenas fornecer informação sobre

$\mathrm{E}(Y_S \mid S = t) = \mathrm{E}(Y_t \mid S = t)$ e $\mathrm{E}(Y_S \mid S = c) = \mathrm{E}(Y_c \mid S = c)$.

No entanto, é claro que $\mathrm{E}(Y_t)$ e $\mathrm{E}(Y_t \mid S = t)$ não são em geral iguais, bem como $\mathrm{E}(Y_c)$ e $\mathrm{E}(Y_c \mid S = c)$. Quando as unidades são designadas aleatoriamente às causas consideradas ($t$ ou $c$), por meio de algum processo físico de randomização (tal como o lançamento de uma moeda), então a determinação de qual causa a unidade $u$ é exposta é considerada estatisticamente independente de qualquer outra variável, incluindo $Y_t$ e $Y_c$. Isso significa que se um processo de randomização é executado, então é plausível considerar $S$ independente de $Y_t$ e $Y_c$ e de todas as outras variáveis sobre $U$. Essa é a hipótese de independência que, quando válida, valida as seguintes equações básicas:

$$\mathrm{E}(Y_t) = \mathrm{E}(Y_t \mid S = t) \quad \text{e} \quad \mathrm{E}(Y_c) = \mathrm{E}(Y_c \mid S = c).$$

Logo,

$$\hat{T} = \mathrm{E}(Y_S \mid S = t) - \mathrm{E}(Y_S \mid S = c) = \mathrm{E}(Y_t \mid S = t) - \mathrm{E}(Y_c \mid S = c) = \mathrm{E}(Y_t) - \mathrm{E}(Y_c) = T$$

Podendo essa quantidade ser estimada a partir dos dados observados por

$$\sum_{u=1}^{U} Y_S(u) I_{(S(u)=t)} U_t - \sum_{u=1}^{U} Y_S(u) I_{(S(u)=c)} U_c \ ,$$

onde $U_t$ e $U_c$ são os números de indivíduos alocados nos grupos tratado e controle, respectivamente, e $I$ é uma variável indicadora de indivíduo tratado ou não.

Assim, o estimador $\hat{T} = \mathrm{E}(Y_t \mid S = t) - \mathrm{E}(Y_c \mid S = c)$, sob a hipótese de independência, é não viciado para $T$, o efeito causal médio, que é o parâmetro de interesse para a solução estatística. Esse estimador é de uso corrente em estudos de comparação de tratamentos e, não raro, não se encontram explicitadas as hipóteses adjacentes que viabilizam sua

avaliação. Holland (1986) chama este estimador de 'efeito causal à primeira vista', enquanto Smith & Sugden (1988) batizaram-no de 'efeito causal aparente'.

Sendo $T$ uma média, goza de todas as características e propriedades inerentes a esta estatística. Assim, se a variabilidade dos efeitos causais $Y_t(u) - Y_c(u)$ sobre todas as unidades $u$ em $U$ é grande, $T$ pode não representar adequadamente o efeito causal de uma determinada unidade $u_0$. Se existe interesse específico em $u_0$, então $T$ pode ser irrelevante independentemente de quão precisa tenha sido sua estimativa.

Uma outra hipótese que torna possível a inferência causal é a de 'efeito constante'. Esta hipótese assume que o efeito de $t$ é o mesmo para todas as unidades. Vale então a seguinte equação:

$$T = Y_t(u) - Y_c(u), \qquad \forall u \in U.$$

Essa hipótese é conhecida pelo nome de 'aditividade' em modelos estatísticos para experimentos, uma vez que o tratamento $t$ 'adiciona' uma quantidade constante $T$ à resposta controle para cada unidade. Sob aditividade, o efeito causal é especialmente importante e fácil de estimar, pois o mesmo efeito vale para cada unidade da população. Em muitos casos em que não vale essa hipótese, já se mostrou que o efeito causal obtido no nível populacional é uma medida de pouco interesse. Por exemplo, em um estudo sobre a eficácia de uma droga nova onde se suspeita de importantes diferenças entre homens e mulheres, os efeitos causais subpopulacionais, para homens e para mulheres, é que são relevantes. Efeitos causais subpopulacionais não são potencialmente mais atraentes apenas porque estão condicionados a características observadas das unidades, mas também porque aditividade provavelmente vale mais neste nível que em um nível populacional (Rubin, 1990a). Eventualmente, havendo interesse em perseguir a validação da hipótese de efeito constante, comumente faz-se uso de transformações adequadas na variável resposta.

A hipótese de efeito constante pode ser parcialmente verificada dividindo-se $U$ em $n$ subpopulações $U_1, U_2, \ldots, U_n$ com base em co-variáveis. Para cada $U_i$, podemos estimar os efeitos causais médios subpopulacionais $T_1$, $T_2$, ... , $T_n$. Uma grande variabilidade entre os $T_i$ é indicação de que a hipótese de efeito constante pode não ser válida. Caso contrário, essa hipótese pode ser plausível.

Existe uma relação entre as hipóteses de homogeneidade e efeito constante. Se vale a hipótese de homogeneidade, vale também a hipótese de efeito constante. Isso porque se, para duas unidades $u_1$ e $u_2$, $Y_t(u_1) = Y_t(u_2)$ e $Y_c(u_1) = Y_c(u_2)$, então

$$Y_t(u_1) - Y_t(u_2) = 0 \text{ e } Y_c(u_1) - Y_c(u_2) = 0 \rightarrow Y_t(u_1) - Y_t(u_2) = Y_c(u_1) - Y_c(u_2) \therefore$$

$$Y_t(u_1) - Y_c(u_1) = Y_t(u_2) - Y_c(u_2) = T.$$

Se faz-se somente a hipótese de efeito constante, não se pode concluir que $\hat{T}$ seja não viciado para o efeito causal médio $T$. Isso porque, sob a hipótese de efeito constante, pode-se escrever $Y_t(u) = T + Y_c(u)$ para todas as unidades. Aplicando-se o operador esperança condicionado a $S = t$ nesta última expressão, tem-se

$$\mathrm{E}(Y_t \mid S = t) = T + \mathrm{E}(Y_c \mid S = t).$$

Mas, $\hat{T} = \mathrm{E}(Y_t \mid S = t) - \mathrm{E}(Y_c \mid S = c)$. Logo,

$$\hat{T} = T + [\mathrm{E}(Y_c \mid S = t) - \mathrm{E}(Y_c \mid S = c)].$$

Sem a hipótese de independência, não há uma razão específica para que o termo entre colchetes acima seja igual a zero, o que seria necessário para que, somente com a hipótese de efeito constante, o estimador $\hat{T}$ pudesse estimar sem viés o parâmetro de interesse $T$.

Já com a hipótese mais forte de homogeneidade de unidades, é fácil mostrar que o efeito causal aparente $(\hat{T})$ é não viesado para o efeito causal médio $(T)$. De fato, sob esta hipótese, tem-se que $Y_k(u_1) = Y_k(u_2)$ para todo par de unidades $u_1$ e $u_2$. Isto é, as respostas potenciais $Y_t(u)$ e $Y_c(u)$ são constantes para qualquer $u \in U$ (porém diferentes para a mesma uni-

dade, a menos que não haja efeito). Assim, $\mathrm{E}(Y_t)$ e $\mathrm{E}(Y_c)$ são também constantes para quaisquer combinações de unidades. Logo,

$$\hat{T} = \mathrm{E}(Y_t \mid S = t) - \mathrm{E}(Y_c \mid S = c) = \mathrm{E}(Y_t) - \mathrm{E}(Y_c) = T \cdot$$

## Hipótese de valor estável unidade-tratamento

Toda a estruturação do Modelo de Rubin, discutida até aqui e sintetizada na Tabela 1, só é adequada sob a condição que Rubin (1980) batizou de 'hipótese de valor estável unidade-tratamento' ou posteriormente, de forma mais simples, de 'hipótese de estabilidade' (Rubin, 1990b). Esta é simplesmente a hipótese *a priori* de que o valor de *Y* para a unidade *u* quando exposta ao tratamento *t* é o mesmo, independentemente de qual mecanismo é usado para designar o tratamento à unidade e de quais tratamentos as outras unidades recebem. Neste último aspecto, coincide com a hipótese de não interferência entre unidades, de Cox (1958).

Embora não seja absolutamente necessária para inferência causal, a hipótese de estabilidade é a hipótese mais simples sob a qual os efeitos causais para cada unidade, tais como $Y_t(u) - Y_c(u)$, podem ser definidos precisamente (Rubin, 1990a). Por ser uma hipótese muito forte, freqüentemente não é válida para muitas situações. Além disso, Dempster (1990) observa que a simplicidade dessa hipótese contrasta com a complexidade dos problemas de causalidade encontrados em situações reais. A hipótese de estabilidade é freqüentemente plausível em experimentos bem planejados. Em outras situações, entretanto, fazem-se necessárias considerações criteriosas antes de se legitimar causalidade com base nos efeitos observados. Quando a hipótese de estabilidade é violada, modelos especiais podem ser construídos a fim de acomodar novas condições, embora tais modelos normalmente assumam hipóteses não testáveis.

As formas mais comuns de violação da hipótese de valor estável unidade-tratamento acontecem quando não existe uniformidade na eficiência

para cada tratamento ou quando existe interferência entre as unidades. Para exemplificar o problema com a uniformidade na eficiência, suponha um estudo no qual se quer avaliar o efeito de uma droga, administrada oralmente, sobre a pressão sangüínea. O grupo tratado, então, é aquele que recebe o comprimido com a substância para a qual se supõe seja hipotensora, e o grupo controle recebe um placebo. Se, para o grupo tratado, os comprimidos variam em eficiência, isto é, se apresentam quantidades variadas da substância considerada e isso pode ter um efeito diferenciado, então o efeito causal $Y_t(u) - Y_c(u)$ não é precisamente definido, uma vez que seu valor depende especificamente de qual comprimido a unidade $u$ recebe. A rigor, para satisfazer a hipótese de valor estável unidade-tratamento, cada comprimido deveria ser considerado um tratamento. Em respostas dicotômicas essa possível violação é menos sentida, na medida em que somente grandes variações nas 'doses' implicariam mudança na resposta. Um exemplo seria a avaliação do efeito do hábito de fumar sobre doença cardiovascular, onde no grupo de fumantes pode haver diferentes níveis de consumo de cigarros. Na maioria das situações práticas, procura-se garantir essa hipótese administrando-se 'doses' ou construindo-se grupos de tratamento tão homogêneos quanto possível. Entretanto, sua verificação normalmente não é viável com base nos dados observados.

A violação da hipótese de estabilidade devido à existência de interferência entre as unidades, por ser mais nitidamente percebida em situações reais, tem sido motivo de maior preocupação. Há interferência entre unidades quando a resposta à causa imputada (tratamento) para uma determinada unidade depende do tratamento designado a outra unidade. Logo, o efeito causal também não é definido, já que seu valor depende de unidades 'vizinhas'. Em epidemiologia, a ocorrência de doenças contagiosas constitui bom exemplo de uma situação em que essa hipótese é claramente violada, já que estas doenças têm como característica fundamental

a presença de cadeias de transmissão de um hospedeiro para outro, de modo que um indivíduo (unidade) eventualmente torna-se infectado dependendo de contatos com outros já infectados.

Como conseqüência da violação da hipótese de estabilidade, faz-se necessária uma expansão da representação das respostas, tal como aquela da Tabela 1, adicionando-se tantas colunas de respostas potenciais quanto necessárias para se considerar a interferência entre as unidades (Rubin, 1990b). Assim, por exemplo, numa situação extremamente simples em que só exista interferência entre as unidades $u_1$ e $u_2$, de tal forma que a resposta da unidade $u_2$ depende do tratamento que a unidade $u_1$ recebe, seriam necessárias quatro colunas para representar as respostas potenciais. Percebe-se imediatamente a dificuldade dessa abordagem para uma situação real envolvendo complexas cadeias de transmissão.

Halloran & Struchiner (1995), estudando a aplicação do Modelo de Rubin em programas de vacinação (intervenção) para doenças contagiosas, observam que a exposição à infecção (interferência) pode também ser vista como um tipo de causa ou manipulação. A exposição à infecção em doenças contagiosas dada pelos outros membros da população, diretamente ou via vetores,[4] é fundamental não só para a transmissão como também para a avaliação dos efeitos da intervenção (vacinação). Entretanto, essa causa deve ser avaliada usando-se um diferente tipo de efeito causal, o efeito indireto. Este é, portanto, aquele efeito devido ao contato com outros indivíduos ou vetores de transmissão, diferente do efeito direto atribuível à vacinação. Assim, os autores discutem uma forma alternativa para solucionar o problema de interferência entre unidades, presente em doenças contagiosas, por meio do condicionamento da análise somente aos indivíduos expostos à infecção.

---

[4] Um vetor pode ser entendido como qualquer ser vivo que veicule o agente infeccioso desde a fonte de infecção (indivíduo, macaco etc.) até o hospedeiro em potencial.

No desenvolvimento de sua análise, definem como $E$ o indicador para exposição à infecção, onde $(E = +)$ representa exposição à infecção e $(E = -)$ representa não exposição à infecção. Os tratamentos (causas) em questão, $t$ e $c$, são receber vacina e não receber vacina, e a resposta $Y$ é também dicotômica, assumindo o valor 0 para não doente e 1 para doente. Considerando a notação já utilizada, as respostas potenciais para uma determinada unidade $u$ quando exposta à infecção seriam, então, $(Y_t(u) \mid E = +)$ e $(Y_c(u) \mid E = +)$. O efeito causal direto condicional para a unidade $u$ é a diferença entre essas duas respostas potenciais. Assim, analogamente a $T$, o efeito causal médio, os autores definem o efeito causal médio condicional à exposição à infecção como

$$T' = \mathrm{E}(Y_t \mid E = +) - \mathrm{E}(Y_c \mid E = +).$$

O termo $\mathrm{E}(Y_c \mid E = +)$ é o quociente entre o número de indivíduos suscetíveis não vacinados que potencialmente tornam-se infectados, se expostos à infecção, e o número total de expostos. Pode, portanto, ser definido como a probabilidade de transmissão esperada em indivíduos suscetíveis não vacinados devido ao contato com indivíduos expostos já infectados. Vale a mesma interpretação para o termo $\mathrm{E}(Y_t \mid E = +)$, substituindo-se não vacinado por vacinado. Se ninguém é exposto à infecção, então $(Y_c \mid E = -)$ e $(Y_t \mid E = -)$ são iguais a zero, de modo que o efeito causal médio na ausência de exposição à infecção será também zero.

Assim, o condicionamento do estudo aos indivíduos expostos à infecção tenta resgatar a hipótese de estabilidade. Isto é, condicionar o estudo aos expostos à infecção torna novamente adequada a representação das respostas potenciais em duas colunas: uma para a resposta ao tratamento 'receber vacina' e outra para a resposta ao controle 'não receber vacina'. Entre as dificuldades dessa solução, Halloran & Struchiner (1995) assinalam que nem todos os indivíduos são expostos à infecção, de modo que a resposta potencial para um determinado indivíduo é dependente não só do mecanis-

mo de designação como também do potencial de exposição à infecção, caracterizando então uma abordagem contrafactual em dois níveis. Isto é, 'se' um indivíduo é exposto à infecção e 'se' é não vacinado, qual a sua resposta? Um outro problema está relacionado à impossibilidade de se distinguir, na prática, dois diferentes tipos de não exposição à infecção. Há aqueles indivíduos que, independentemente do tratamento recebido, não serão expostos à infecção, e há aqueles que não serão expostos sob particulares designações de tratamento. Por fim, dado que pessoas não expostas à infecção não são incluídas na análise, o número de indivíduos no estudo poderia ser muito reduzido, sobretudo se muitos expostos à infecção são vacinados.

Assim, inferência causal em doenças contagiosas, onde não vale a hipótese de estabilidade devido à presença de interferência entre as unidades, tem como dificuldade principal a identificação das fontes de exposição à infecção, comumente não possível em situações reais, de modo que se exigem hipóteses não testáveis específicas sobre as situações de exposição à infecção, tais como picadas de mosquitos, relações sexuais ou o número de contatos com indivíduos já infectados. Alternativamente, o estudo de intervenções em doenças contagiosas pode ser conduzido, substituindo-se as informações sobre a exposição à infecção por hipóteses parcialmente não testáveis que modelem a dinâmica da transmissão do agente infeccioso.

## O Mecanismo de Designação de Tratamentos

Diferentemente de um modelo determinístico de causa que procura 'demonstrar' causalidade, tal como o modelo de Rothman, um modelo estatístico ou probabilístico procura 'inferir' causalidade. Esse modelo não necessariamente nega a visão de que a ocorrência de uma doença poderia ser perfeitamente determinada por certa quantidade de fatores, apesar de muitos

deles não serem conhecidos ou eventualmente não observados de forma apropriada. Assim, a incerteza presente nos modelos estatísticos serve sobretudo para expressar a ignorância sobre o processo causal e sobre como observá-lo na prática, e não porque se acredita ser a ocorrência da doença um fenômeno aleatório. O Modelo de Rubin, não obstante apresentar uma estrutura determinística, tem a sua discussão inserida em um contexto probabilístico em razão, principalmente, do Problema Fundamental da Inferência Causal.

É por meio do mecanismo que designa qual tratamento as unidades recebem que se identifica a natureza probabilística necessária para se tentar inferir causalidade. Mais precisamente, qualquer método estatístico para inferência causal exige especificações para os procedimentos que geram os dados observados. Isto é, faz-se necessário conhecer as estruturas probabilísticas de registro de dados por meio dos mecanismos de amostragem e de designação de tratamentos. Aqui, com o objetivo de explorar mais detalhadamente a aplicabilidade da idéia de causa contida no Modelo de Rubin, não será considerado nenhum esquema de amostragem. Assim, a aleatoriedade presente será toda devida ao mecanismo de designação de tratamentos, assumindo-se, portanto, que todas as $U$ unidades participam do estudo. Entretanto, tem-se visto que o próprio mecanismo de designação de tratamentos pode refletir tanto o esquema de seleção das unidades de uma população finita quanto a indicação de qual tratamento cada unidade recebe (Rubin, 1978). Smith & Sugden (1988) separam estes dois mecanismos e diferenciam a análise pela ordem com que eles se sucedem: amostrando-se as unidades antes de designar o tratamento ou, alternativamente, designando-se os tratamentos às unidades e posteriormente amostrando-se unidades de cada estrato de tratamento.

Desse modo, para que se possa inferir causalidade, independentemente do procedimento estatístico a ser utilizado, é necessária a explicitação de um mecanismo de designação de tratamentos – um modelo probabilístico para o

processo pelo qual se obtêm os valores de *Y* que de fato são observados. Em outras palavras, é necessário especificar o processo que conduz aos valores observados de *S*. Comumente, para a situação dicotômica considerada, tem-se na prática imaginado as unidades recebendo um determinado tratamento, *t* por exemplo, com probabilidade $p(X)$ e, conseqüentemente, o tratamento alternativo com probabilidade $1-p(X)$, onde $p(X)$ é uma função desconhecida das co-variáveis *X*. Assim, o mecanismo de designação de tratamentos pode ser escrito genericamente por meio da seguinte expressão probabilística, atribuindo-se os valores 0 e 1 para os tratamentos *c* e *t*, respectivamente:

$$\Pr(\mathbf{S} \mid \mathbf{X}) = \prod_{u=1}^{U} p(X_u)^{S(u)} [1 - p(X_u)]^{1-S(u)},$$

onde as variáveis em negrito indicam vetor ou matriz de dados para todas as unidades.

Essa expressão revela o mecanismo de designação como dependente de co-variáveis observáveis, a matriz $\mathbf{X}_{U \times q}$. Entretanto, um mecanismo de designação pode também ser dependente de variáveis não observáveis ou somente parcialmente observáveis, a matriz *U*x2 de respostas potenciais $\mathbf{Y} = (\mathbf{Y}_t, \mathbf{Y}_c)$. Seja *W* o conjunto de todas as outras co-variáveis pré-exposição, que diferentemente das co-variáveis *X*, são desconhecidas ou não passíveis de observação. Desse modo, o mecanismo de designação de tratamentos deve ser concebido de forma geral como dependente dessas três quantidades aleatórias. Simbolicamente, $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}, \mathbf{W})$.

Um mecanismo é dito 'controlado' quando está livre da dependência de **Y** e **W** (Smith & Sugden, 1988). Sem considerar a dependência sobre **W**, Rubin (1990a, 1991) classifica alguns mecanismos de designação de tratamentos de acordo com suas propriedades. Assim, uma classe muito importante é aquela constituída dos mecanismos ditos 'não confundidos' (o neologismo é inevitável). Estes são os mecanismos que independem da variável resposta **Y**, isto é, $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} \mid \mathbf{X})$ para todos os valores possíveis

de **S**, **X** e **Y**. A dependência de um mecanismo sobre **X** pode usualmente ser tratada diretamente na análise, enquanto a dependência sobre **Y**, em especial sobre sua parte não observada, em geral é difícil de ser considerada de maneira satisfatória. Daí a importância dessa classe de mecanismos.

Uma importante característica da especificação de um mecanismo de designação é que ele deve ser 'probabilístico', no sentido de que cada unidade tenha uma probabilidade de designação dos tratamentos $t$ ou $c$ satisfazendo a desigualdade

$$0 < \Pr(S(u) = t \mid \mathbf{X}, \mathbf{Y}) < 1, \qquad \forall u \in U.$$

Se alguma unidade não tem chance de ser designada para algum tratamento, é sensato considerar o porquê e se a inclusão de tal unidade é de fato adequada ao estudo planejado (Rubin, 1991).

Um mecanismo de designação é dito 'ignorável' para os valores observados de $X$, $Y$ e $S$ se

$$\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}_t, \mathbf{Y}_c) = \Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}_S).$$

Isto é, o mecanismo pode ser escrito como dependente somente dos dados observados.

Sob a condição de ser 'não confundido' e 'probabilístico', isto é, se

$$(Y_t, Y_c) \perp S \mid X \quad \text{e} \quad 0 < \Pr(S(u) = t \mid X) < 1,$$

onde o símbolo $\perp$ significa independência, o mecanismo é conhecido como 'fortemente ignorável'. Observa-se que se o mecanismo é 'fortemente ignorável', é também 'ignorável', mas a recíproca não é verdadeira (Rosembaun & Rubin, 1983). Essa classe de mecanismos de designação de tratamentos também está dentro da classe dos mecanismos ditos 'controlados'.

Para inferência causal, o mecanismo de designação de tratamentos assume uma função vital. Pode-se dizer que o ceticismo que acompanha a atribuição de causa com base em estudos observacionais surge das potenciais dificuldades de se colocar um mecanismo de designação aceitável. Isto é, a principal fonte de incerteza na análise de um estudo observacional não é

geralmente o método de inferência empregado, mas sim a especificação do mecanismo de designação (Rubin, 1991). Embora na maioria das vezes esteja completamente ausente, este tipo de discussão deve ser um importante componente da análise de qualquer estudo observacional para efeitos causais.

A importância do mecanismo de designação para inferência causal pode ser ilustrada pela seguinte situação hipotética, baseada em um exemplo de Smith & Sugden (1988). Considere um estudo observacional do tipo coorte, em que indivíduos suscetíveis a determinada doença são acompanhados por certo período de tempo e alocados nos diferentes níveis de exposição (causas) sem qualquer controle sobre este mecanismo. Os dados observados sobre casos incidentes de hepatite A diante das duas causas alternativas postuladas, presença ou não de água tratada no domicílio, para as três mil crianças existentes em certa região e acompanhadas por um ano estão sumariados na Tabela 2.

Tabela 2 – Casos de hepatite A em um ano, segundo condição da água do domicílio (dados hipotéticos)

| Tratamentos ou causas | Total acompanhado | Resposta ($Y_S$) | |
|---|---|---|---|
| | | Casos (1) | Não casos (0) |
| Com água tratada ($t$) | 499 | 165 | 334 |
| Sem água tratada ($c$) | 2.501 | 577 | 1.924 |
| TOTAL | 3.000 | 742 | 2.258 |

Os dados mostram um risco significativamente maior para aqueles que dispõem de água tratada no domicílio (165/499 = 0,3307) em relação àqueles que não possuem (577/2501 = 0,2307), com um efeito causal aparente ($\hat{T}$) de 0,3307 - 0,2307 = 0,1. Como já dito, em epidemiologia, tal medida é conhecida como 'risco atribuível'. O risco atribuível é usado para quantificar o risco da doença no grupo exposto (no exemplo, com água

tratada) que pode ser considerado atribuível à exposição, uma vez que remove o risco de doença presente de forma geral devido a outras causas (o risco no grupo não exposto). Assim, a interpretação do risco atribuível é dependente da suposição de que existe uma relação de causa e efeito entre exposição e doença.

Uma conclusão precipitada com base no risco atribuível obtido com base nos dados da Tabela 2 indicaria que possuir água tratada no domicílio, ao contrário de um conhecimento já consolidado, é fator causal para hepatite A. O número de casos a mais de hepatite A entre aqueles que possuem água tratada atribuível a esta condição é de 1 para cada 10 crianças. Entretanto, o quão aproximada essa estimativa está do efeito causal verdadeiro, desprezada a variação amostral, pode depender de co-variáveis observáveis ou não e também da própria resposta. Como já observado, em um modelo estatístico de causalidade, essa dependência é tratada por meio do mecanismo de designação de tratamentos. Inferir causalidade com base no efeito causal aparente sem considerar essas possíveis dependências pode levar a conclusões enganadoras, pois este efeito pode ser bem diferente do efeito causal verdadeiro. Mesmo mecanismos de designação de tratamentos controlados, que dependem apenas de co-variáveis observáveis, podem produzir grandes diferenças. Se uma co-variável $X$ é tal que para valores grandes desta variável mais provavelmente se observa a resposta 1 para o tratamento $t$ e a resposta 0 para o tratamento $c$, e se a alocação dos tratamentos é obtida inadvertidamente sem a consideração de $X$, então os efeitos causais aparente e verdadeiro serão bem distantes. O condicionamento do mecanismo de designação em $X$ produziria um efeito causal aparente ajustado, de modo que o efeito causal verdadeiro poderia ser estimado sem viés.

Para o exemplo considerado, a matriz de respostas potenciais $\mathbf{Y}_{U\text{x}2} = (\mathbf{Y}_t, \mathbf{Y}_c)$ é uma matriz onde cada $Y_k(u)$, $k = \{t,c\}$ e $u = 1, 2,...,U$ assume o valor 0 ou 1. Assim, os $U$ pares de respostas potenciais

$(Y_t(u), Y_c(u))$ podem tomar os valores (0,0), (0,1), (1,0) e (1,1). Se as freqüências desses pares são hipoteticamente conhecidas e dadas pela Tabela 3, a seguir, então o efeito causal médio verdadeiro seria

$$T = \mathrm{E}(Y_t) - \mathrm{E}(Y_c) = \frac{300+100}{3000} - \frac{600+100}{3000} = \frac{300-600}{3000} = -0{,}1 .$$

Tabela 3 – Exemplo de um mecanismo de designação não ignorável e freqüência hipotética dos pares de respostas potenciais

| Respostas potenciais $(Y_t, Y_c)$ | Freqüência | Pr(*S(u)=t* ) |
|---|---|---|
| (0,0) | 2.000 | 0,11 |
| (0,1) | 600 | 0,19 |
| (1,0) | 300 | 0,52 |
| (1,1) | 100 | 0,09 |
| TOTAL | 3.000 | |

Suponha que o mecanismo não seja ignorável e que a probabilidade de designação do tratamento *t* a uma unidade dependa do par de respostas potenciais, como indicado pela mesma Tabela 3. Sob este mecanismo, as freqüências de casos e não casos de hepatite A pelos níveis de exposição considerados que se espera observar podem ser calculadas. A resposta 1 (caso) é observada, ou quando *t* é aplicado com probabilidade 0,52 ao par de respostas potenciais (1,0), cuja freqüência é de 300 crianças, ou com probabilidade 0,09 ao par (1,1), que tem freqüência 100. Isto é, a freqüência esperada de casos sob o tratamento *t* é (300x0,52)+(100x0,09)=165. Da mesma forma, a freqüência de não casos sob o tratamento *c* (sem água tratada), que se esperaria observar se a verdade fosse aquela representada pela Tabela 3, seria de 2000x(1-0,11)+300x(1-0,52)=1924. As outras duas freqüências esperadas, casos sob o tratamento *c* e não casos sob o tratamento *t*, podem ser obtidas de forma análoga, cujos resultados são 577 e 334, respectivamente.

É curioso observar que as freqüências esperadas coincidem exatamente com os valores observados pelo estudo e apresentados na Tabela 2.

A finalidade desse exemplo é ilustrar quão perigosa pode ser a interpretação de resultados se baseada em estudos observacionais. Se, eventualmente, por um alguma razão o mecanismo que alocou as unidades aos tratamentos não é ignorável e este fato não é do conhecimento do analista, as estimativas resultantes sofrerão erros sistemáticos, podendo inclusive gerar uma medida de efeito de igual magnitude mas de sentido contrário, como é o caso desse exemplo, perdendo, conseqüentemente, sua interpretação causal.

O mecanismo de designação de tratamentos é o elemento, então, que assume papel dominante em uma formulação estatística de causalidade. Dependendo do tipo de estudo empregado, mecanismos particulares podem ser usados a fim de subsidiar conclusões de causalidade. Classicamente, randomização em estudos experimentais (Rubin, 1974) e, mais recentemente, o escore de propensão em estudos observacionais (Rosembaun & Rubin, 1983) têm sido advogados como mecanismos apropriados para se inferir causalidade.

## Randomização

Randomização é um tema cuja literatura disponível é bastante extensa e data o início do século XX, com os trabalhos de Fisher. Apesar disso, a discussão de suas características e propriedades continua sendo extremamente atual. Formalmente, randomização proporciona um mecanismo com base no qual é possível a obtenção das propriedades probabilísticas das estimativas. As principais propriedades são:

- o efeito causal aparente $\hat{T}$ é uma estimativa não viesada de T, o efeito causal verdadeiro. Esta propriedade já fora observada quando se comentou sobre a hipótese de independência;

- declarações probabilísticas bem definidas podem ser feitas para indicar quão incomuns valores observados de $\hat{T}$ seriam sob um específico efeito causal hipotetizado. P-valores, intervalos de confiança e razões de verossimilhança para parâmetros causais são computados sob a hipótese de que as indicações alternativas de tratamento são respostas igualmente prováveis. Esta hipótese é perfeitamente justificada sob randomização.

Greenland (1990) tem dito que inferir estatisticamente causalidade com base em intervalos de confiança, p-valores ou razões de verossimilhança é de pouco significado se o mecanismo de designação à exposição é desconhecido ou sabidamente não randomizado. Um mecanismo randomizado proporciona, então, o elo entre inferência estatística e parâmetros causais. De forma análoga, amostras aleatórias fornecem a chave para o relacionamento entre inferência estatística e parâmetros populacionais.

Um importante fato enfatizado por Rubin (1978) é que randomização, quando empregada, garante independência *a priori* entre o mecanismo de designação de tratamentos e diversos fatores ou atributos das unidades, incluindo aqueles eventualmente também causais. Assim, randomização funciona como um fiador da validade da representação matemática de medidas de efeito de fatores causais observados, permitindo modelar os efeitos de fatores não observados como erros aleatórios (Dempster, 1990).

A propriedade que tem o estimador considerado – no caso de se adotar um mecanismo randomizado – de não ser tendencioso não deve trazer muito conforto para o epidemiologista. O conceito estatístico de viés e o conceito epidemiológico de confundimento, embora possam ser relacionados, são bem diferentes. Viés estatístico se refere a uma média ponderada diferente de zero das estimativas dos efeitos causais, com pesos dados pela distribuição de probabilidade de tais estimativas. Confundimento é visto como falta de comparabilidade dos grupos considerados, sendo, por-

tanto, característica de uma particular alocação. Porém, no caso artificial de se quantificar o grau de confundimento de cada possível alocação como a diferença entre a estimativa e o parâmetro causal de interesse, pode-se concluir que a não tendenciosidade de experimentos randomizados corresponde a um confundimento médio igual a zero sobre a distribuição dos resultados das possíveis alocações (Greenland, 1990). O epidemiologista, ao tentar interpretar o resultado de um estudo, está interessado no grau de confundimento de uma particular estimativa observada, cuja magnitude não é diretamente obtida por um estudo randomizado. Entretanto, sob randomização pode-se fazer a probabilidade de um confundimento grave, tal como interpretado anteriormente, tão pequena quanto se queira, aumentando-se o tamanho das coortes sob os tratamentos considerados. No entanto, esse recurso, quase sempre uma grande dificuldade em situações reais, não é uma garantia absoluta de que não há confundimento severo em um particular estudo. É ainda possível que um 'infeliz' resultado tenha ocorrido e, assim, o grau de confundimento seja grande. Desse modo, se as coortes são grandes e não há evidências de falta de comparabilidade entre elas, o processo de randomização proporciona um mecanismo em que se deposita alta credibilidade na hipótese de que as coortes são aproximadamente comparáveis, desde que não haja nenhuma violação grosseira do protocolo de designação dos tratamentos às unidades.

Ainda dentro dessa lógica, atualmente faz-se uma interpretação mais atraente de um ensaio randomizado. Supõe-se que randomização garanta que os grupos sob as condições de exposição consideradas são comparáveis, onde o sentido de comparabilidade é que o resultado observado para o grupo exposto teria sido o mesmo se o grupo não exposto tivesse sido exposto, e vice-versa. Isso porque, como já dito aqui, supõe-se assegurada a hipótese de que as respostas são independentes da designação dos tratamentos e de que co-variáveis não observadas são balanceadas nos dois grupos. Em ou-

tras palavras, sob randomização, não importa qual dos dois grupos de indivíduos será exposto ao fator causal considerado.

Entretanto, dependendo de características específicas de um estudo, nem mesmo a randomização garante estimativas não tendenciosas do efeito biológico de interesse. Struchiner & Halloran (1996), estudando os efeitos de vacinas para doenças contagiosas, mostraram como a exposição à infecção pode ser um fator de confundimento para o efeito causal de interesse mesmo em ensaios randomizados. Exposição à infecção, tal como picada de mosquito ou contato sexual, é vista como uma co-variável freqüentemente não registrada nos estudos que se dá exclusivamente pela ação da natureza (Halloran & Struchiner, 1995).

Sintetizando, um mecanismo de designação construído com base em algum processo randomizado, tal como o lançamento de uma moeda, onde a indicação de um tratamento se daria pela ocorrência de cara e, obviamente, a ocorrência de coroa implicaria indicação do tratamento alternativo, pode ser escrito como

$$\Pr(\mathbf{S}) = \left(\frac{1}{2}\right)^{U},$$

sendo as $2^U$ possíveis indicações igualmente prováveis, proporcionando portanto um mecanismo independente das variáveis $X$ e $Y$ e, mais interessante que estas, também de $W$, as co-variáveis não observadas. Daí a credibilidade dada a esse mecanismo quando o interesse é atribuir causa. Para o Modelo de Rubin, a validade desta última expressão é suficiente para inferência causal.

## Escore de propensão

Como visto, sob randomização, os dois grupos sob tratamento podem ser diretamente comparáveis porque provavelmente suas unidades são similares. Sem esse mecanismo, comparações diretas podem ser enganadoras,

pois as unidades expostas a um tratamento podem, em geral, diferir sistematicamente das unidades expostas ao outro tratamento. Nessas condições, escores balanceadores podem ser usados para que a comparação entre os grupos considerados seja reveladora do efeito que se quer estimar.

Um escore balanceador, $b(X)$, é uma função das co-variáveis observadas *X*, tal que a distribuição condicional de *X* dado $b(X)$ é a mesma para o grupo tratado ( $S = t$ ) e para o grupo controle ( $S = c$ ). Isto é, dado $b(X)$, *S* e *X* são independentes. Simbolicamente,

$$S \perp X \mid b(X).$$

O escore balanceador mais trivial é obviamente $b(X) = X$, e outros de maior interesse são funções mais complexas de *X*.

Rosenbaun & Rubin (1983) mostraram que se um mecanismo de designação é fortemente ignorável dado *X*, então a diferença entre as médias dos grupos tratamento e controle para cada valor de um escore balanceador é uma estimativa não viesada do efeito do tratamento naquele valor e, conseqüentemente, pareamento ou estratificação sobre este escore devem produzir estimativas não viesadas do efeito causal médio verdadeiro.

Assim, um outro mecanismo de designação, com indicação para utilização em estudos observacionais, é obtido com base em um particular escore balanceador conhecido como 'escore de propensão'. Este é definido como a probabilidade condicional de designação de um particular tratamento dado um vetor de co-variáveis observadas. O interessante desse mecanismo é que o ajustamento para um escore de propensão escalar é suficiente para remover vieses, em razão de todas as co-variáveis observadas. A seleção de unidades tratadas e controles com o mesmo valor para esse escore terão a mesma distribuição de *X*.

Em experimentos randomizados, o escore de propensão é uma função conhecida, havendo portanto uma especificação apropriada para $\Pr(S = t \mid X)$. No entanto, em estudos observacionais, o escore de propen-

são é quase sempre uma função desconhecida, de modo que uma especificação para ele deve ser estimada com base em dados observados por meio de algum modelo adequado, como, por exemplo, o modelo logístico. Para um bayesiano, essas estimativas são probabilidades preditoras *a posteriori* de designação do tratamento $t$ para uma unidade com vetor $X$ de co-variáveis. Sob um modelo logístico, fazendo $t = 1$ e $c = 0$, o escore de propensão $\Pr(S = t \mid X)$ é dado por

$$\Pr(S = t \mid X) = \frac{e^{\beta_0 + \sum_{i=1}^{q} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{q} \beta_i x_i}}$$

onde os $\beta_i$ são coeficientes estimados com base nos dados observados $S$ e $X$.

Já foi utilizada a notação para a probabilidade de uma determinada unidade receber um tratamento, $t$ por exemplo, como $p(X)$. Assim, $p(X) = \Pr(S = t \mid X)$ e, atribuindo-se novamente os valores 1 e 0 para os tratamentos $t$ e $c$, respectivamente, tem-se o seguinte mecanismo de designação, também já apresentado, construído com base no escore de propensão:

$$\Pr(\mathbf{S} \mid \mathbf{X}) = \prod_{u=1}^{U} p(X_u)^{S(u)} [1 - p(X_u)]^{1 - S(u)}$$

Uma interessante utilização do escore de propensão acontece quando se criam pares de unidades tratamento-controle. Para cada unidade que recebe um tratamento, busca-se outra para ser seu controle que tenha um escore de propensão aproximadamente igual. Com isso, inferências sobre o efeito causal devido ao tratamento podem ser obtidas livres dos efeitos das variáveis usadas para construir o escore. Esse tipo de uso do escore de propensão em estudos observacionais é análogo à randomização em experimentos controlados pareados. Dentro de cada par, tratado e controle têm aproximadamente a mesma probabilidade preditora, dado $X$, de ser tratado. Randomização em um experimento de comparação pareado é mais interes-

sante, é claro, na medida em que sustenta a hipótese de não dependência de *X* bem como de *W*, o conjunto de co-variáveis não observadas.

Como ilustração das potencialidades desse mecanismo, considere-se a análise de causalidade desenvolvida por Rubin (1991) em que se busca estimar os possíveis efeitos em virtude da entrada no mercado de uma droga psiquiátrica com seu nome genérico em concorrência a seu nome comercial. O grupo tratado ($S=t$) foi constituído por aqueles pacientes que optaram pela troca, enquanto o grupo controle ($S=c$) consistiu daqueles que mantiveram o uso da droga com seu nome comercial. Foram observadas 53 co-variáveis pré-intervenção entre variáveis demográficas e relacionadas ao uso da droga, cujo controle pensou-se ser conveniente para que os efeitos pudessem ser medidos e atribuídos à causa em questão e não a uma possível diferença nestas co-variáveis entre os grupos. Obteve-se assim o escore de propensão $\Pr(S = t \mid X)$ para cada unidade, por meio de um modelo logístico para as co-variáveis *X* ou transformações delas. Com base nesse escore para cada unidade tratada, buscou-se uma unidade controle com o valor mais próximo. A conseqüência desse processo de pareamento foi sintetizada em uma tabela de freqüências, tal como a Tabela 4. Esta tabela conta a freqüência, por faixas de magnitude, de uma medida de viés padronizada para as 53 co-variáveis envolvidas, ou eventualmente utilizando-se transformações apropriadas, antes e depois do pareamento. Essa medida de viés foi obtida por meio da seguinte expressão:

$$\frac{\bar{x}_d}{\sqrt{\frac{(s_t^2 + s_c^2)}{2}}}$$

onde $\bar{x}_d$ foi a diferença entre as médias dos grupos tratado e controle para cada variável (ou transformação), antes e após o pareamento, e $s_t^2$ e $s_c^2$ foram as variâncias de cada grupo antes do pareamento.

Nota-se que antes do pareamento havia duas co-variáveis com viés acima de 0,25 e 10 (quase 19%) com viés de pelo menos 0,10. Após o pareamento, houve grande concentração das variáveis com medida de viés inferior a 0,05, o que levou Rubin (1991) a considerar o procedimento bem efetivo. Isto é, os dois grupos sob investigação puderam ser considerados homogêneos, pelo menos com relação às co-variáveis observadas.

Tabela 4 – Distribuição de freqüência de uma medida de viés padronizada para 53 co-variáveis antes e após pareamento

| Viés Padronizado | Antes N | Antes % | Após N | Após % |
|---|---|---|---|---|
| Até 0,05 | 30 | 56,6 | 49 | 92,5 |
| 0,05 - 0,10 | 13 | 24,5 | 4 | 7,5 |
| 0,10 - 0,15 | 3 | 5,7 | 0 | 0,0 |
| 0,15 - 0,20 | 5 | 9,4 | 0 | 0,0 |
| 0,20 - 0,25 | 0 | 0,0 | 0 | 0,0 |
| 0,25 ou + | 2 | 3,8 | 0 | 0,0 |
| TOTAL | 53 | 100,0 | 53 | 100,0 |

Admitindo-se esse procedimento como adequado, pode-se adotar um mecanismo de designação mais simplificado. Se, entre as $U$ unidades presentes no estudo, encontramos $M \leq U/2$ pares, de tal forma que as unidades de cada par tratamento-controle apresentem a mesma probabilidade de serem tratadas, então o mecanismo de designação reduz-se a

$$\Pr(\mathbf{S}) = \left(\frac{1}{2}\right)^M,$$

se as unidades de cada par recebem tratamentos diferentes. Assim, cada um dos $2^M$ possíveis valores de **S** que designam $t$ a uma unidade do par e $c$ à outra, são igualmente prováveis, e qualquer **S** que designe as duas unidades do par o mesmo tratamento tem probabilidade 0.

Resumindo, as limitações e propriedades de um mecanismo construído com base no escore de propensão são bem conhecidas. Pareamento sobre esse escore balanceia as co-variáveis observadas *X*. Entretanto, diferentemente da randomização, não balanceia co-variáveis não observadas, exceto aquelas correlacionas com *X*. Outra questão importante é que ajustamento sobre esse escore balanceia *X* apenas em média. Em um particular estudo, mesmo após um ajustamento, é possível ainda encontrar por mero acaso algumas variáveis não balanceadas entre os grupos sob investigação. Daí a necessidade, quase sempre presente, de se estabelecerem hipóteses adicionais, eventualmente não testáveis, a fim de validar uma atribuição de causalidade.

A abordagem mais comumente encontrada para controle de co-variáveis eventualmente confundidoras tem se dado por meio de análises multivariadas, incorporando a noção de modelagem estatística. No entanto, a contribuição dos modelos estatísticos para inferência causal, com os quais se tenta descrever as medidas de efeito causal de interesse por meio dos coeficientes do modelo, tem sido apresentada e criticada em muitos trabalhos (Greenland, 1979; Robins & Greenland, 1986; Vandenbrouke, 1987; Greenland, 1989b). Alternativamente, apesar de só mais recentemente se encontrarem na literatura clínica e epidemiológica trabalhos que utilizem o escore de propensão como um instrumento para controle simultâneo de muitas co-variáveis observadas, Rosembaun e Rubin têm, em dupla ou individualmente, defendido em vários trabalhos este mecanismo como apropriado para a discussão de causalidade com base em estudos observacionais. D'Agostino (1998) apresenta um excelente tutorial desse procedimento.

# Inferência Estatística

Como já mencionado, o objetivo, quando se analisam os dados disponíveis, é obter alguma informação sobre os pontos de interrogação ilustrados na Tabela 1 e, assim, poder estimar efeitos causais tais como $T$ para as unidades sob investigação. Repetindo, as interrogações se referem aos valores de $Y$ que seriam observados se, contrariamente ao fato, as unidades tivessem recebido o tratamento alternativo àquele que receberam. Entretanto, é objetivo também estender a informação desses efeitos para unidades em que nenhum dos elementos do par de respostas potenciais $(Y_t, Y_c)$ tenha sido observado, tais como as outras unidades da população das quais se supõe ter extraído aquelas que de fato estão sob análise, ou mesmo para unidades que futuramente possam ser expostas aos tratamentos. Em outras palavras, quer-se 'inferir' estatisticamente efeitos causais. Rubin (1990a, 1991) tem assinalado e comparado quatro métodos de inferência estatística, apontando a diferença entre eles como conseqüência da forma com que usam os dados observados para obter informação sobre os valores não observados, os quais, como se tem visto, são necessários para que se definam os efeitos causais. Mais importante que as diferentes formas de construção desses métodos é que todos compartilham a mesma estrutura conceitual de causalidade presente no Modelo de Rubin.

Os quatro métodos considerados por Rubin (1990a) como conceitualmente distintos são:

- teste de hipóteses nulas pontuais sob um mecanismo randomizado;
- inferência de parâmetros causais baseada em possíveis alocações aleatórias de tratamentos;
- inferência bayesiana para efeitos causais;
- inferência freqüentista.

Apesar do Problema Fundamental da Inferência Causal, talvez a demonstração mais simples de que é possível a estimação de efeitos causais seja possível por meio do estabelecimento de uma hipótese nula pontual tal como

$$H_0\text{: } Y_t(u) = Y_c(u), \qquad \forall u \in U.$$

sob essa hipótese, **Y**, a matriz de respostas potenciais $U$x2 é de fato completamente revelada, uma vez que ou $Y_t(u)$ ou $Y_c(u)$ é observado, e $Y_t(u) = Y_c(u)$. Sendo $Y_s$ o valor observado de $Y$ e definindo-se $\mathbf{Y}_{H_0}$ como o valor revelado sob $H_0$, tem-se $\mathbf{Y}_{H_0} = (\mathbf{Y}_S; \mathbf{Y}_S)$. Dessa forma, são especificados valores conhecidos para cada ponto de interrogação da Tabela 1. Em outras palavras, $H_0$ em conjunto com os dados observados implica valor específico para cada interrogação e, assim, todos os valores e todos os efeitos causais passam a ser conhecidos.

Segundo o desenvolvimento de Rubin (1990a), dois fatos importantes se seguem. Primeiro, os valores das probabilidades das possíveis designações de tratamentos, especificada pelo mecanismo de designação $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y})$, podem ser obtidos não só para a particular designação $S$ observada como também para qualquer outra designação possível, dados $\mathbf{X} = \mathbf{X}_{\text{obs}}$ e $\mathbf{Y} = \mathbf{Y}_{H_0}$. Em segundo lugar, os valores do vetor $\mathbf{Y}_{Ux1} = \{Y_{S(u)}(u)\}$ podem ser encontrados para cada possível $S$ sob $H_0$. Em particular, $\mathbf{Y}_{H_0} = \mathbf{Y}_S$. Assim, o valor de qualquer estatística $Z = Z(S, X_{\text{obs}}, Y_{H_0})$ pode ser calculada não apenas para a específica designação de tratamentos observada $S$, mas também para todas as outras possíveis. Como resultado, escolhendo-se uma direção para valores incomuns de $Z$, vale o seguinte cálculo:

$$\Pr(Z \text{ mais extremo que } Z_{\text{obs}} \mid \mathbf{X} = \mathbf{X}_{\text{obs}}, \mathbf{Y} = \mathbf{Y}_{H_0}) =$$

$$= \sum_S \Pr(S \mid \mathbf{X} = \mathbf{X}_{\text{obs}}, \mathbf{Y} = \mathbf{Y}_{H_0}) \text{ x } \delta(Z \text{ mais extremo que } Z_{\text{obs}}),$$

onde $d$ é uma função indicadora, assumindo o valor 1 quando o argumento for verdadeiro e 0 quando for falso. Essa expressão fornece o nível de

significância (ou p-valor) para $H_0$ baseado na estatística de teste $Z$. Sob um mecanismo probabilístico não confundido corresponde ao teste randomizado de Fisher. Estes p-valores podem também ser calculados com mecanismo confundidos ou mesmo não probabilísticos, embora a interpretação dos resultados seja pouco interessante (Rubin, 1990a).

Já se viu que se o p-valor é pequeno, então ou um evento raro aconteceu, ou a hipótese nula não é verdadeira, ou ainda o mecanismo de designação colocado está errado. Dessa maneira, a qualquer hipótese nula pontual pode ser atribuído um p-valor, onde pontual significa que $H_0$ em conjunto com os valores observados $Y_s$, implica o conhecimento de todos os valores de $Y$. Esse método de inferência tem como grande vantagem sua simplicidade e imparcialidade, que são particularmente interessantes quando $Z$ e sua direção de raridade são definidos antes que qualquer dado seja observado.

Outro aspecto interessante desse método se refere ao rico debate entre Neyman e Fisher, cuja origem pode ser atribuída à especificação da hipótese nula. Enquanto Fisher defendia uma hipótese nula tal como a apresentada acima, ou seja, de ausência de efeito em cada unidade, Neyman propunha como hipótese nula a igualdade entre os efeitos médios populacionais, isto é, $H_0$: $\mathrm{E}(Y_t) = \mathrm{E}(Y_c)$. Rubin (1990b) discute maiores detalhes sobre as conseqüências da abordagem de Neyman.

Estreitamente relacionado a esse método por também funcionar sob aleatoriedade, distingue-se outro método para se inferir causalidade em que, entretanto, hipóteses nulas pontuais não assumem mais um papel central. A abordagem se concentra na inferência de parâmetros causais sob a distribuição de um estimador causal obtida com base nas possíveis e aleatórias designações dos tratamentos. Por ser mais associada à idéia de levantamentos amostrais aleatórios, Rubin (1990a) assinala que essa bordagem deve ser vista mais como uma avaliação da qualidade de procedimentos de

inferência propostos, tais como erro quadrático médio e poder, do que propriamente um método para inferência por si.

Inicialmente define-se um parâmetro causal, $T = \mathrm{E}(Y_t - Y_c)$ por exemplo, podendo ser inclusive função de *X*, o que caracterizaria um parâmetro causal subpopulacional. Em seguida, busca-se um estimador $\hat{T} = \hat{T}(S, X, Y)$ tal que seja aproximadamente não tendencioso para *T*, isto é,

$$\mathrm{E}(\hat{T} \mid X, Y) \cong T,$$

onde a aleatoriedade é devida a *S* com distribuição dada por $\Pr(S \mid X, Y)$. Em geral, a esperança de uma estatística depende de **(X,Y)**. Assim, a determinação de $\hat{T}$ que satisfaça a esperança acima é fácil com mecanismos de designação probabilísticos não confundidos, tal como randomização, mas em mecanismos diferentes pode ser difícil ou mesmo impossível sem a definição de fortes hipóteses *a priori* (Rubin, 1990a). É necessária, ainda, a determinação de um estimador $\hat{V} = \hat{V}(S, X, Y)$ para a variância de $\hat{T}$ que seja aproximadamente não tendencioso e que tenha variabilidade menor que a de $\hat{T}$. Assim, assumindo-se normalidade da distribuição de $T - \hat{T}$, inferências para *T* são obtidas a partir da declaração probabilística

$$(T - \hat{T}) \sim \mathrm{N}(0; \hat{V}).$$

A escolha dos estimadores é geralmente baseada em princípios tais como variância mínima, não tendenciosidade e menor erro quadrático médio. Apesar de reconhecer que esse método proporciona resultados coincidentes com os do método anterior, Rubin (1990a) aponta a dificuldade dessa abordagem em se adequar a problemas mais complicados e reais, como por exemplo distribuições não normais. Resumindo, esse método de inferência trata **X** e **Y** como fixados mas **Y** como desconhecido, e obtém as características dos estimadores sob todas as possíveis designações de tratamentos dadas pelo mecanismo $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y})$.

O método bayesiano proporciona abordagem mais direta para se inferirem efeitos causais, na medida em que se concentra na distribuição

*a posteriori* de *T*. Essa é, pois, a distribuição do parâmetro causal de interesse condicionada a valores observados e especificações de modelos probabilísticos para as variáveis envolvidas. Isto é, inferência bayesiana de efeitos causais é obtida com base na colocação de uma distribuição de probabilidade *a priori* para **Y** dado **X** e da especificação probabilística do mecanismo de designação de tratamentos. Enquanto no método de hipóteses nulas pontuais os pontos de interrogação da Tabela 1 eram especificados com base nos dados observados em conjunto com $H_0$, no método bayesiano busca-se a distribuição deles dados todos os valores observados. Busca-se, portanto, a distribuição *a posteriori* de $\mathbf{Y}_{\bar{S}}$, a componente não observada de **Y**. A distribuição de $\mathbf{Y}_{\bar{S}}$ pode ser escrita como

$$\Pr(\mathbf{Y}_{\bar{S}} \mid \mathbf{X}, \mathbf{S}, \mathbf{Y}_S) = \frac{\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) \Pr(\mathbf{Y} \mid \mathbf{X})}{\int \Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) \Pr(\mathbf{Y} \mid \mathbf{X}) \, d\mathbf{Y}_{\bar{S}}}$$

onde $\mathbf{Y} = (\mathbf{Y}_S, \mathbf{Y}_{\bar{S}})$ 'particiona' **Y** em valores observados e não observados. E, no caso de um mecanismo de designação de tratamentos não confundido, isto é, se

$$\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} \mid \mathbf{X}),$$

então a distribuição de $\mathbf{Y}_{\bar{S}}$ é simplificada, sendo dada por

$$\Pr(\mathbf{Y}_{\bar{S}} \mid \mathbf{X}, \mathbf{Y}_S) = \frac{\Pr(\mathbf{Y} \mid \mathbf{X})}{\int \Pr(\mathbf{Y} \mid \mathbf{X}) \, d\mathbf{Y}_{\bar{S}}}.$$

Os valores observados de $(\mathbf{X}, \mathbf{S}, \mathbf{Y}_S)$ em conjunto com a distribuição *a posteriori* de $\mathbf{Y}_{\bar{S}}$ implica distribuição para qualquer função de $(\mathbf{X}, \mathbf{S}, \mathbf{Y})$ tais como os efeitos causais médios populacional ou subpopulacional, ou mesmo efeitos causais individuais. Assim, dentro de uma estrutura bayesiana, inferências causais podem ser obtidas por exemplo a partir de $\Pr(T \mid \mathbf{X}, \mathbf{S}, \mathbf{Y}_S)$. Próprio desse método é a necessidade de se estabelecer um modelo para a distribuição de **Y** dado **X**. Usualmente apela-se para o teorema construído por de Finetti e escreve-se

$$\Pr(\mathbf{Y} \mid \mathbf{X}) = \int \prod_{u=1}^{U} f(Y_u \mid X_u, \theta) \Pr(\theta) d\theta$$

onde $Y_u$ é a *u*-ésima linha de **Y** e $f(Y_u \mid X_u, \theta)$ é o modelo comum para a distribuição condicional de **Y** dado **X** para cada linha de **(Y,X)** dado o parâmetro não observável $\theta$ com distribuição *a priori* Pr($\theta$). Não faz muito tempo, a grande dificuldade dessa abordagem se concentrava na identificação de distribuições conjugadas, de modo a se obter um procedimento analiticamente tratável. Entretanto, essa barreira tem sido vencida com a incorporação de procedimentos de simulação.

O quarto método de inferência citado por Rubin (1990a; 1991) domina a prática. O método freqüentista é similar ao método de inferências de parâmetros causais na sua formulação, mas em vez de suas características serem obtidas sob possíveis **S** do mecanismo de designação $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y})$ para um fixado **(X,Y)**, são obtidas sob repetidas observações de **Y** a partir da distribuição amostral $\Pr(\mathbf{Y} \mid \mathbf{X}; \theta)$, dada por

$$\Pr(\mathbf{Y} \mid \mathbf{X}; \theta) = \prod_{u=1}^{U} f(Y_u \mid X_u; \theta)$$

com $\theta$ constante mas desconhecido. Essa abordagem proporciona resultados estatísticos matemáticos bem conhecidos para $\theta$, especialmente quando *f* segue uma distribuição normal.

Rubin (1991) critica esse método porque, apesar de formular quase todas as hipóteses que o método bayesiano formula, não permite uma especificação *a priori* para $\theta$. Do ponto de vista prático, uma especificação para Pr($\theta$) é muito menos sensível do que qualquer especificação para $f(Y_u \mid X_u, \theta)$, principalmente em estudos observacionais. Assim, por exemplo, enquanto na teoria bayesiana intervalos de confiança para $\theta$ têm uma interpretação probabilística imediata, dentro do método freqüentista esta interpretação depende do argumento condicional de que amostras repetidas seriam necessárias para verificar a freqüência relativa de intervalos que de fato compreendiam $\theta$.

Uma outra forma, mais genérica, de se apresentar conceitualmente a questão da inferência é encontrada em Smith & Sugden (1988), escrevendo-se a função de verossimilhança das variáveis envolvidas. O processo de inferência causal envolve basicamente a consideração de três conjuntos de variáveis: as respostas $Y$, as co-variáveis conhecidas $X$ e as desconhecidas $W$. Formalmente, pode-se escrever parametricamente de forma conveniente a distribuição conjunta dessas variáveis como:

$$f(\mathbf{W} \mid \mathbf{Y}, \mathbf{X}; \psi) g(\mathbf{Y} \mid \mathbf{X}; \theta) h(\mathbf{X}; \phi),$$

onde se assume que os parâmetros $\psi$, $\theta$ e $\phi$ são distintos. Tomando-se também de forma geral o mecanismo de designação como $\Pr(S \mid \mathbf{X}, \mathbf{Y}, \mathbf{W})$ e admitindo-se que todas as unidades estão presentes no estudo, a função de verossimilhança pode ser escrita como

$$h(\mathbf{X}; \phi) \int_W \int_{\bar{S}} \Pr(S \mid \mathbf{Y}, \mathbf{X}, \mathbf{W}) f(\mathbf{W} \mid \mathbf{Y}, \mathbf{X}; \psi) g(\mathbf{Y} \mid \mathbf{X}; \theta) \; d\mathbf{Y} d\mathbf{W}.$$

E as condições que implicariam um mecanismo que sustentasse uma interpretação causal seriam:

$$S \perp \mathbf{Y} \mid \mathbf{X}, \mathbf{W} \quad \text{e} \quad \mathbf{Y} \perp \mathbf{W} \mid \mathbf{X}.$$

Diante dessas considerações sobre os métodos de inferência estatística, percebe-se que os recursos que a estatística tradicionalmente utiliza para fazer inferências sobre associações podem ser transferidos para o caso em que se estuda causalidade com base na adoção explícita de um modelo especialmente planejado para este fim.

# 3

# Questões Epidemiológicas Relacionadas

Até aqui se tem visto como a estatística pode contribuir para a discussão sobre inferência causal, procurando explorar as propriedades e limitações de um modelo estatístico de causalidade devido a Rubin. É interessante, também, discutir as interfaces dessa contribuição com a teoria e a prática epidemiológica, abordando algumas questões metodológicas relacionadas.

A importância do instrumental estatístico no desenvolvimento da teoria epidemiológica tem cada vez mais ficado evidente. Embora construída sobre o tripé clínica, medicina social e estatística, a epidemiologia tem experimentado nos últimos anos um *boom* de reconhecimento graças, em parte, à utilização e ao aperfeiçoamento de técnicas quantitativas avançadas especialmente desenvolvidas para responder à complexidade dos problemas por ela enfrentados. Assim, a incorporação de um modelo estatístico de causalidade parece ser relevante por ser esta, a causalidade, o grande desafio da investigação epidemiológica.

Pode-se entender inferência epidemiológica como o processo de obtenção de inferências, tais como a predição de ocorrências de doenças ou a identificação de suas causas, com base em dados epidemiológicos, isto é, dados relacionados à ocorrência de doenças em populações. Essas inferências devem poder ser feitas sem os benefícios diretos de evidências experimentais e também sem a necessidade de uma teoria estabelecida sobre a etiologia

da doença, ou seja, sem o estabelecimento de um mecanismo causal. Cabe aqui um parêntese, já que o termo 'população' em epidemiologia merece atenção especial. Apesar de os epidemiologistas observarem indivíduos, as interpretações dos resultados são baseadas na combinação de dados de muitas unidades. Razões para o uso de populações em epidemiologia estão relacionadas a dois grandes objetivos: avaliar programas de melhora do *status* de saúde de grupos específicos e, mais interessante que isso, permitir que uma investigação faça inferências causais, usando métodos estatísticos, sobre as relações entre certas exposições e o *status* de saúde.

Como exemplo de que inferências epidemiológicas devem ser possíveis mesmo sem os benefícios de um estudo experimental, considere como problema a verificação da hipótese de que cafeína é fator causal para doença cardiovascular. Essa hipótese pode ser testada por meio da estimação do risco atribuível a esse fator, que, como já visto, é obtido pela diferença entre o risco de doença cardiovascular entre indivíduos usuários dessa substância e o risco entre indivíduos não usuários. Entretanto, diferentemente de uma exposição experimental, o uso de cafeína é de opção pessoal e tem sido apontada sua correlação com diversas outras preferências, principalmente o fumo. Portanto, mesmo que cafeína não tenha nenhum efeito sobre o risco de doença cardiovascular, não se esperaria encontrar uma equivalência de riscos entre usuários e não usuários de cafeína. Assim, a análise deve ser conduzida necessariamente através de um estudo observacional que seja capaz de corrigir eventuais diferenças entre os grupos considerados. É essa tentativa de correção que tem estimulado muito não só o aparecimento de novas técnicas estatísticas como também o estabelecimento de critérios de diagnóstico e controle dessas técnicas, de modo a validar ou não as hipóteses a elas adjacentes.

Um desafio dos estudos epidemiológicos é, portanto, a estimação isolada do risco atribuível (ou de qualquer outra medida de efeito pertinente)

a uma determinada exposição – no exemplo, uso de cafeína –, livre dos efeitos devidos a outros fatores, de modo que as eventuais conclusões de causalidade sejam válidas. No entanto, nesse momento surge uma outra característica fundamental da pesquisa epidemiológica. Eventualmente, a estimação independente de uma medida de efeito de determinado fator pode ser de pouca valia se existe um outro fator que age sinergicamente com ele. Existe sinergismo quando a presença de um fator modifica o efeito biológico do outro e um caso particular acontece quando os indivíduos têm a doença somente se expostos aos dois fatores, mas não a um deles isoladamente. Sinergismo – conceito da biologia – é, na prática, comumente avaliado por meio do conceito estatístico de interação. Diz-se que existe interação estatística entre dois fatores, A e B, se é necessário um parâmetro adicional para descrever adequadamente o risco de doença em virtude da exposição conjunta a esses fatores. Em oposição, dois fatores podem ser considerados estatisticamente independentes se o risco de doença, dada uma exposição conjunta a esses fatores, pode ser adequadamente modelado como uma função dos efeitos separados, onde esta função pode ser aditiva, multiplicativa ou, ainda, assumir outra forma. Assim, a determinação se existe ou não interação estatística depende do modelo estatístico adotado (Siemiatycki & Thomas, 1981).

Desse modo, ao lado do confundimento, a interação de fatores, entendida, portanto, como a interdependência entre dois fatores que alteram a magnitude de um dado efeito, é uma outra importante preocupação metodológica. Entretanto, o interesse principal em uma discussão de causalidade presente nos estudos epidemiológicos é o julgamento da validade da associação observada entre exposição e doença.

# O Conceito de Validade

Validade pode ser entendida como a capacidade de um particular estudo epidemiológico encontrar critérios lógicos que apontem para ausência de vieses. Assim, um estudo de associação entre uma exposição e uma doença é considerado válido, e com conseqüente interpretação causal, se está livre de vieses que possam, alternativamente à exposição proposta, explicar seus resultados.

Os vieses possíveis em um particular estudo podem ser devidos a várias fontes, e a maior parte deles é comum a qualquer desenho epidemiológico. Tradicionalmente, como já dito aqui, os vieses têm sido classificados como viés de informação, viés de seleção e confundimento. Greenland (1991) propõe uma classificação mais abrangente, que, embora não tenha ecoado na literatura, parece representar uma estrutura mais conveniente para discutir a validade dos estudos e o processo de inferência epidemiológica. Em sua abordagem, qualquer estudo epidemiológico cujo objetivo seja a avaliação de causalidade deve considerar as seguintes condições: validade de comparação, validade de seguimento, validade de especificação e validade de mensuração. Em estudos de caso-controle, especificamente, além de serem mais sujeitos a fontes adicionais de erros de mensuração, deve-se considerar também a validade de seleção de casos e controles. Percebe-se imediatamente a complexidade que envolve a validade de determinada medida de efeito, visto que é necessário o cumprimento de uma série de condições suficientes. Essas quatro condições são vistas como condições de validade 'internas', pois se referem apenas à estimação de efeitos dentro dos grupos de exposição considerados, sem preocupação quanto à generalização dos resultados para outras populações. São condições suficientes para validade, mas não necessárias, no sentido de que uma eventual violação pode não produzir viés.

Pode-se acreditar que haja validade de comparação se as unidades que de fato foram expostas, caso não o tivessem sido, teriam apresentado aproximadamente a mesma distribuição para a variável resposta (proporção de doença) que as unidades não expostas. Vale o mesmo raciocínio para as unidades não expostas: tivessem elas sido expostas, se esperaria observar o mesmo resultado que o das unidades de fato expostas. Validade de comparação simplesmente significa que a distribuição da variável resposta para as unidades de um determinado grupo de tratamento (exposição) prediz o que teria acontecido com o grupo de tratamento alternativo, tivessem as unidades deste último grupo experimentado a condição do outro. Ou ainda, uma outra forma de dizer isso é afirmar que os grupos tratado (exposto) e controle (não exposto) são 'comparáveis' ou 'permutáveis' com respeito à variável resposta, independentemente da situação de exposição que eventualmente cada unidade experimente. Falta de validade de comparação deve resultar em uma estimativa viesada da medida de efeito causal considerada e diz-se que há 'confundimento' na estimativa. Como se trata da principal preocupação da investigação de causa, essa questão é retomada adiante em um item específico, no qual é discutida mais detalhadamente, procurando-se interfaces entre a forma com que a epidemiologia trata a questão e o Modelo de Rubin.

Validade de seguimento acontece quando, dentro de cada grupo de tratamento considerado (tratado e controle), o risco de censura[5] não está associado ao risco de doença. Desse modo, dada a validade de seguimento, espera-se que o risco de doença dentro de cada grupo em um determinado momento seja o mesmo tanto para unidades perdidas (ou retiradas do estudo por razões concorrentes) até este momento quanto para unidades acompanhadas além deste momento. No exemplo do uso de cafeína como fator

[5] Censura diz respeito à finalização do período de seguimento por uma razão diferente daquela que se esperaria observar, no caso, ocorrência da doença em questão.

causal para doença cardiovascular, poder-se-ia imaginar um viés de seguimento, já que tabagismo, que está associado à exposição em questão, relaciona-se também a uma taxa maior de mortalidade por doença cardiovascular e também por outras doenças. Ou seja, indivíduos expostos à cafeína estariam potencialmente mais sujeitos à censura – por causas de morte relacionadas ao tabagismo – do que os indivíduos não expostos.

Quando se discutiram os mecanismos de designação de tratamentos, particularmente o escore de propensão, verificou-se a necessidade de se estabelecer um modelo estatístico. Mais genericamente, as técnicas estatísticas assumem algum tipo de modelo para o processo de geração dos dados a serem analisados. E, na ausência de randomização ou amostras aleatórias, raramente será possível identificar um modelo absolutamente correto. Logo, algum tipo de erro de especificação deve ser esperado. No caso de respostas dicotômicas, o modelo amostral tradicionalmente empregado é o binomial. Um modelo amostral deve ser entendido como uma expressão matemática que descreve a probabilidade de observação dos possíveis dados como função de parâmetros, quase sempre desconhecidos. Em um estudo de coorte, por exemplo, a validade de um modelo binomial depende da hipótese de que os riscos dentro de cada grupo de exposição permanecem constantes por todo o período de acompanhamento. Mais especificamente, o risco de doença cardiovascular não é constante ao longo do tempo para as coortes acompanhadas uma vez que, à medida que os indivíduos vão envelhecendo, tal risco vai aumentando. Logo, a utilização do modelo binomial, nesse exemplo, não gozaria de validade de especificação.

Além da fonte de erro devido ao modelo amostral, falta de validade de especificação pode ser devida também a um modelo estrutural, isto é, a uma expressão matemática que descreve os parâmetros do modelo amostral como função de co-variáveis, tal como um modelo logístico. Associado a qualquer modelo estrutural encontra-se sempre um conjunto de hipóteses que para

serem validadas são comumente confrontadas com os dados observados. Na literatura, é comum encontrar o modelo estrutural incorporado ao modelo amostral, sendo essa combinação designada genericamente como 'modelo estatístico'. Assim, de uma estimativa de uma medida de efeito pode-se dizer que tem validade de especificação se é obtida com base em um modelo estatístico correto, ou pelo menos aproximadamente correto. E se o modelo amostral ou o modelo estrutural usado na análise for incorreto, as estimativas resultantes podem estar viesadas. Como comentário adicional, mesmo quando um eventual erro de especificação não conduza a um viés, testes estatísticos ou intervalos de confiança para as estimativas ficam comprometidos (Greenland, 1991).

Já foi ressaltado que a contribuição da estatística para o processo de inferir causalidade passa necessariamente pela mensuração de variáveis que permitam a quantificação dos efeitos causais. Desse modo, medir adequadamente as variáveis envolvidas é condição *sine qua non* para uma possível interpretação de causa. Entretanto, do ponto de vista prático ou operacional, qualquer estudo está sujeito a erros de mensuração por várias fontes, o que obviamente pode contribuir para um possível viés na estimativa. Assim, pode-se dizer que uma estimativa de uma determinada medida de efeito goza de validade de mensuração se ela não sofre de vieses em conseqüência de erros de mensuração nas variáveis envolvidas no estudo. Para ajudar na tarefa de identificar possíveis erros de mensuração, pode ser útil classificá-los de acordo com suas fontes. E erros de uma determinada fonte podem ser ainda classificados de acordo com características que são preditoras da direção do viés que eles produzem.

Greenland (1991) propõe um esquema de classificação para os erros de mensuração bastante abrangente, dividindo-os quanto à fonte, quanto ao tipo e quanto à dependência entre eles. Quanto à fonte, os erros podem ser classificados como erros de procedimento – tais como instrumentos mal

calibrados ou falta de lembrança em estudos retrospectivos, erros de uma variável *proxy*, ou seja, erros por se utilizar uma variável substituta à variável de interesse – e erros de construção, que surgem de ambigüidades na definição das variáveis. Independentemente da fonte, os erros podem ser classificados como diferenciais ou não diferenciais, se a direção ou a magnitude dos erros tende a variar ou não com os valores verdadeiros de outras variáveis, respectivamente. Finalmente, diz-se que erros na mensuração de duas variáveis são dependentes se a magnitude ou a direção do erro cometido na mensuração de uma variável está associada à magnitude ou à direção do erro cometido em uma outra variável. Se não existe associação entre os erros, diz-se que são independentes. Questão importante suscitada por essa classificação é uma maior 'valorização' da validade de mensuração. Se suspeita-se que os 'instrumentos' utilizados para medir as variáveis envolvidas não medem adequadamente aquilo que se pensa que estejam medindo, então pouca credibilidade deveria ser dada a qualquer resultado eventualmente obtido. Por exemplo, conceitos epidemiológicos tais como 'sensibilidade' e 'especificidade' dizem respeito ao poder, ou às probabilidades, que tem um teste diagnóstico de classificar como doente ou não doente indivíduos que de fato estejam doentes ou não doentes, respectivamente. Logo, conhecer a magnitude dessas probabilidades seria necessário para, pelo menos, identificar o sentido do erro.

Diante dessas considerações, em que se registra uma série de condicionantes acerca da validade de um estudo, parece justificável a insegurança sempre presente quando o assunto é estabelecimento de causa. O conhecimento dos possíveis vieses envolvidos em um particular estudo é, portanto, fundamental para que, sempre que possível, possam ser verificados a partir dos dados disponíveis. Fazendo uma associação com a teoria contida no Modelo de Rubin, a discussão de alguns possíveis vieses fica embutida nas hipóteses relacionadas ao modelo, que precisariam, portanto, ser testadas. De novo, nem sempre as hipóteses adjacentes podem ser postas à prova.

# Confundimento e Permutabilidade

Confundimento tem sido reconhecido como um dos principais problemas da pesquisa epidemiológica e já foi aqui citado diversas vezes. Em razão disso, muito se tem escrito e já há bastante tempo sobre esse problema metodológico, registrando-se um rico debate não só sobre abordagens alternativas para seu controle, mas também sobre o seu próprio conceito. Do ponto de vista epidemiológico, vê-se confundimento como um problema surgido a partir de uma diferença intrínseca aos riscos de doença entre as populações exposta e não exposta (Greenland & Robins, 1986). Isso significa que a diferença existiria mesmo se a exposição estivesse ausente em ambas as populações. Retomando o exemplo em que a exposição sob consideração é o uso de cafeína, esperar-se-ia encontrar riscos para doença cardiovascular diferentes entre usuários e não usuários de cafeína, mesmo se nenhum daqueles indivíduos usuários tivesse de fato sido exposto à cafeína. Essa diferença poderia ser explicada pela falta de comparabilidade, ou confundimento, devida à associação entre cafeína e tabagismo. O conceito de confundimento está baseado, portanto, na idéia de comparabilidade.

Entretanto, parece que o próprio conceito do problema ainda é motivo de discussão, sendo o termo usado comumente para se referir a três problemas distintos. A forma mais antiga se relaciona a um tipo de viés na estimação de um efeito causal, utilizado na literatura epidemiológica e sociológica. Um conceito mais recente associa confundimento à idéia de 'não colapsibilidade', isto é, uma certa medida de efeito bruta é diferente daquela obtida pela combinação desta medida com estratos de uma terceira variável. Um terceiro conceito está relacionado à impossibilidade de separação do efeito principal e o efeito de interação, presente comumente em uma análise de variância. É difícil distinguir adequadamente esses conceitos. Em particular, os conceitos de confundimento como um viés na estimação do efeito e como 'não

colapsibilidade' são freqüentemente tratados como idênticos. Greenland et al. (1999) abordam essa questão e explicitam essas diferenças. Independentemente do conceito, o importante é que, para se fazer inferência causal, é necessária a garantia de comparabilidade entre os grupos considerados.

Greenland (1989a), ao comentar o trabalho de Wickramaratne & Holford (1987) sobre confundimento em estudos epidemiológicos, identifica três ambientes apropriados para essa discussão. O primeiro corresponde exatamente ao Modelo de Rubin, o qual, segundo ele, é a melhor formalização atualmente disponível para se tratar do problema de confundimento. Como foi possível observar aqui, essa abordagem explora a noção de 'não ignorabilidade' da designação dos tratamentos. Um segundo ambiente, imaginado por Greenland & Robins (1986), trata confundimento sob o conceito de permutabilidade. Essa questão é retomada adiante por apresentar forte ligação com o Modelo de Rubin. A terceira abordagem, devida a Gail (1986), trata confundimento sob a noção de ajustamento para co-variáveis balanceadas.

Greenland & Robins (1986) discutem o conceito de confundimento epidemiológico baseados em um modelo desenvolvido para se estudar efeitos individuais. Sob esse modelo, percebem confundimento como um problema de não identificabilidade de parâmetros. Isto é, valores distintos de um parâmetro de interesse desconhecido podem gerar a mesma distribuição de dados observados. Associam também a esse conceito a idéia de permutabilidade, cujo significado é aquele próprio à teoria bayesiana, qual seja: são esperados os mesmos resultados caso as condições de exposição dos grupos sejam permutadas (Migon & Gamerman, 1999). Esse aspecto corresponde, na realidade, à hipótese de homogeneidade desenvolvida no Modelo de Rubin. Aliás, como poderá ser observado, este modelo de efeitos individuais assemelha-se bastante ao Modelo de Rubin, visto, porém, sob uma ótica exclusivamente epidemiológica.

Fazendo-se uma reprodução simplificada de Greenland & Robins (1986), considere-se a situação na qual se deseja estudar o efeito de um fator de exposição dicotômico sobre o risco de uma doença em um período de tempo a risco especificado. Existem quatro tipos de indivíduos de acordo com suas possíveis respostas aos tratamentos considerados. Assim, um indivíduo classificado como tipo 1 é aquele que se torna doente independentemente da exposição, e o tipo 4 é aquele imune à doença. Os indivíduos tipo 2 ou 3 são os suscetíveis. A Tabela 5 ilustra esse fato, onde o valor 1 é indicativo de presença de doença e o valor 0, de ausência. Após observar um único indivíduo do grupo tratado sobre o período de tempo a risco e verificar que ele contraiu a doença, não é possível dizer, sem uma informação adicional, se ele é do tipo 1 – aquele indivíduo condenado a adoecer sob qualquer condição – ou do tipo 2, aquele indivíduo suscetível à doença pela exposição. Da mesma forma, se ele não tivesse contraído a doença também não seria possível dizer a qual tipo pertenceria: ao 3 ou ao 4. Ou seja, independentemente do que é observado, não é possível dizer se a exposição tem efeito. Esse é um problema de não identificabilidade.

Tabela 5 – Tipos de indivíduos quanto a suas possíveis respostas (1 ou 0) a um tratamento dicotômico

| Tipo e descrição dos indivíduos | Tratamento | | Tipo de efeito |
|---|---|---|---|
| | Tratado | Controle | |
| Tipo 1: Indivíduo condenado a adoecer | 1 | 1 | Sem efeito |
| Tipo 2: Indivíduo suscetível | 1 | 0 | Causal |
| Tipo 3: Indivíduo suscetível | 0 | 1 | Preventivo |
| Tipo 4: Indivíduo imune à doença | 0 | 0 | Sem efeito |

Nota: um efeito preventivo é também um efeito causal.

Se, adicionalmente ao indivíduo exposto, observa-se um indivíduo não exposto, quatro possíveis resultados se apresentam:

a) ambos os indivíduos adoecem;

b) somente o indivíduo exposto adoece;

c) somente o não exposto adoece;

d) nenhum dos dois adoece.

Pela mesma razão exposta, não é possível saber se a exposição tem efeito. Entretanto, se é razoável supor que os dois indivíduos são do mesmo tipo, então torna-se possível deduzir se a exposição tem ou não algum efeito. Isso porque, na hipótese de que os indivíduos são do mesmo tipo, as respostas (a) e (d) significam que a exposição não tem efeito. Do mesmo modo, as respostas (b) e (c), quando combinadas com a hipótese, implicam causalidade. Assim, a combinação da resposta observada com a hipótese de equivalência dos indivíduos implica identificabilidade do efeito. Essa equivalência pode ser colocada em termos de permutabilidade dos indivíduos. Se as condições de exposição dos dois indivíduos tivessem sido trocadas (permutadas), o mesmo resultado seria obtido.

Para completar a apresentação da noção de confundimento por meio do modelo de efeitos individuais de Greenland & Robins (1986), é necessária a consideração de uma população de indivíduos. Considere uma coorte de $U_t$ indivíduos expostos a um tratamento *t*, inicialmente livres de doença, acompanhados ao longo de um período a risco. Seja $p_j$, $j = 1, 2, 3$ e 4, a proporção de indivíduos dessa coorte que são do tipo *j*, de acordo com a Tabela 5. Considere também uma coorte de $U_c$ indivíduos controles (ou não expostos) e equivalentes proporções de indivíduos do tipo *j* dadas por $q_j$. Ao final do tempo considerado, os resultados poderiam ser resumidos em uma tabela 2x2 tal como a Tabela 6, a seguir.

Tabela 6 – Freqüências observadas de doentes, como função dos tipos de indivíduos, segundo coortes a risco

| Grupos de exposição | Total acompanhado | Freqüência | |
|---|---|---|---|
| | | Doentes | Não Doentes |
| Tratado ($t$) | $U_t$ | $D_t = (p_1+p_2)\text{x}U_t$ | $ND_t = (p_3+p_4)\text{x}U_t$ |
| Controle ($c$) | $U_c$ | $D_c = (q_1+q_3)\text{x}U_c$ | $ND_c = (q_2+q_4)\text{x}U_c$ |

Para atribuição de causa, seria necessário conhecer no grupo tratado as proporções indicadoras de existência de efeito, ou seja, os valores $p_2$ e $p_3$, embutidos nas freqüências observadas $D_t$ e $ND_t$. O risco atribuível à exposição ($\hat{T}$), obtido pela diferença entre os riscos nas populações expostas ($\hat{R}_t$) e não expostas ($\hat{R}_c$) e que tem representado a medida de efeito causal de interesse, é dado por

$$\hat{T} = \hat{R}_t - \hat{R}_c = \frac{D_t}{U_t} - \frac{D_c}{U_c}$$

Só com o valor calculado de $\hat{T}$ não é possível dizer se efetivamente há algum efeito, pois para qualquer valor de $\hat{T}$, positivo, negativo ou zero, não se pode identificar as proporções $p_2$ ou $p_3$. Fazendo-se, entretanto, a hipótese de comparabilidade entre as coortes, isto é, assumindo-se que a proporção de indivíduos que adoeceriam se a exposição estivesse ausente é a mesma para ambas as coortes, seria obtida apenas uma identificabilidade parcial. Isso porque, sob essa condição, vale a equação $q_1 + q_3 = p_1 + p_3$. Logo,

$$\hat{T} = \frac{D_t}{U_t} - \frac{D_c}{U_c} = (p_1 + p_2) - (q_1 + q_3) = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3$$

Então, se $\hat{T}>0$ implica que $p_2$ também maior que zero, o que significa que no estudo houve alguns indivíduos que tiveram a doença 'por causa' da exposição. Da mesma forma, $\hat{T}<0$ significa que alguns tiveram a doença 'prevenida' pela exposição, isto é, $p_3>0$. Porém, se $\hat{T}=0$, só se pode deduzir que $p_2 = p_3$. Para se concluir que não houve efeito, e assim atingir completa

identificabilidade, é necessária a hipótese adicional de que $p_3$ (ou $p_2$, alternativamente) seja igual a zero, isto é, assumir que a exposição nunca previne. Assim, sob essa hipótese adicional, $\hat{T}=0$ implica $p_2=0$, indicando portanto não ter havido efeito. A hipótese de que $q_1 + q_3 = p_1 + p_3$ pode ser vista também como uma hipótese de permutabilidade parcial, e dir-se-ia que se as condições de exposição fossem trocadas, o valor observado do risco na ausência da exposição teria sido o mesmo. Para uma completa permutabilidade seria necessária a hipótese adicional de que $q_1 + q_2 = p_1 + p_2$, ou seja, existiria a mesma relação entre exposição e risco se as condições de exposição fossem permutadas.

A finalidade desta seção foi apresentar a discussão de confundimento tal como ela tem estado presente em um contexto epidemiológico. Diante das características desse modelo determinístico de efeitos individuais, usado por Greenland & Robins (1986) para discutir confundimento, uma analogia com o Modelo de Rubin parece inevitável. Nesse nível individual, a falta de identificabilidade dos efeitos, remediada pela hipótese de permutabilidade, corresponde ao Problema Fundamental da Inferência Causal, cuja solução correspondente seria a hipótese de homogeneidade. Também pode ser percebida, embutida nos quatro tipos de indivíduos citados, a noção de respostas potenciais, própria do Modelo de Rubin. Cada combinação possível de exposição e resposta potencial no Modelo de Rubin corresponderia a um dos quatro tipos de indivíduos no modelo de efeitos individuais.

Do ponto de vista da formulação do Modelo de Rubin, a discussão sobre confundimento está concentrada no mecanismo de designação de tratamentos. E, como já visto, existe uma classe de mecanismos ditos 'não confundidos', que são aqueles que não dependem da variável resposta **Y**, isto é, $\Pr(\mathbf{S} \mid \mathbf{X}, \mathbf{Y}) = \Pr(\mathbf{S} \mid \mathbf{X})$ para todos os valores possíveis de **S**, **X** e **Y**, sendo a dependência sobre **X** passível de ser controlada na medida em que são covariáveis observadas. Se adiciona-se a condição $0 < \Pr(S(u) = t \mid X) < 1$,

viu-se que o mecanismo é conhecido como sendo 'fortemente ignorável'. Assim, o Modelo de Rubin trata o problema de confundimento por meio da idéia de 'ignorabilidade'.

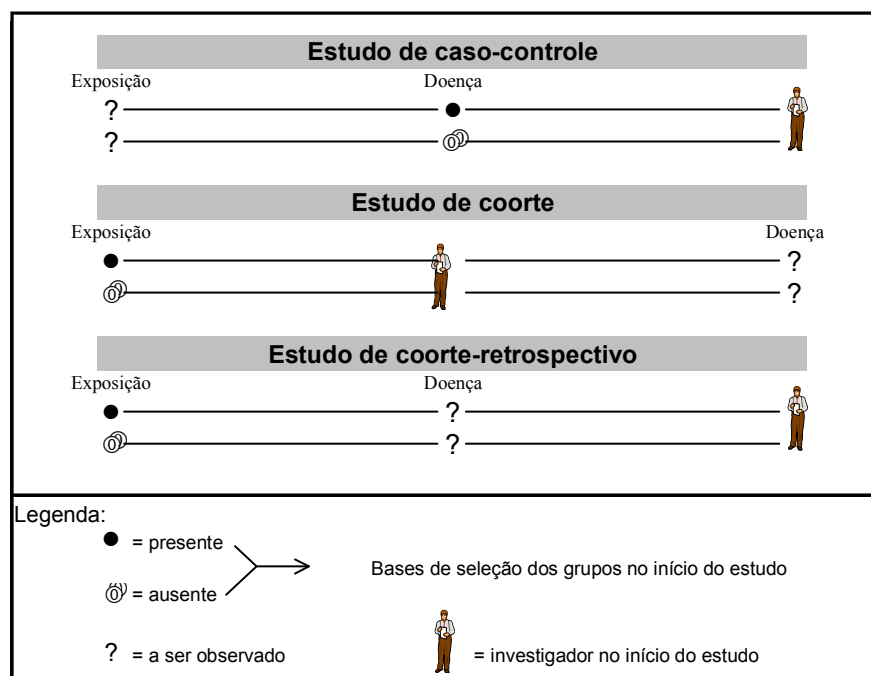
# Os Estudos Epidemiológicos

De forma geral, os estudos epidemiológicos podem ser divididos entre aqueles experimentais e aqueles observacionais, dependendo do controle de que o investigador dispõe sobre o mecanismo que designa a condição de exposição das unidades. Entre os observacionais, destacam-se os estudos de coorte, os estudos de caso-controle e os estudos de coorte-retrospectivo, de acordo com a forma com que são coletadas as unidades e o *timing* de observação.[6] A Figura 2, adiante, ilustra esse conceito. Embora a epidemiologia disponha de outros tipos de estudos, tais como os estudos ecológicos, ou ainda particulares desenhos dentro dos citados, essa classificação é suficiente para se explorar de forma geral a noção de causa contida nos estudos epidemiológicos.

Em um estudo experimental, que no ambiente epidemiológico é conhecido como um ensaio clínico ou um estudo de intervenção, a designação dos tratamentos às unidades é controlada pelo experimentador. Isto é, os tratamentos são designados para as unidades com base em um mecanismo casual, comumente um gerador de números aleatórios, que é de conhecimento do experimentador, de modo que só o acaso determina quem recebe qual tratamento. Sob essa condição, espera-se, portanto, que as unidades recebendo os diferentes tratamentos sejam comparáveis, tal como discutido no capítulo sobre randomização. Entretanto, o acaso poderia em princípio fazer os grupos tratado e controle diferirem de maneira significativa, não

[6] *Timing* refere-se à relação cronológica entre as observações dos *status* de doença e exposição e suas ocorrências de fato.

sendo difícil quantificar o seu impacto potencial e distingui-lo do efeito do tratamento. Testes estatísticos usuais e intervalos de confiança fazem precisamente isso. Sob esse desenho, portanto, toda a teoria desenvolvida para o Modelo de Rubin se aplica diretamente. A causa a ser avaliada bem como a causa alternativa são estabelecidas *a priori* e, uma vez alocadas aleatoriamente nas *U* unidades investigadas, qualquer medida de efeito pode ser obtida diretamente, com a crença geral de que irá satisfazer aos critérios de validade discutidos – salvo pelo acaso e por particulares situações –, tal como comentado na seção sobre randomização.

Figura 2 – *Timing* dos principais estudos epidemiológicos observacionais em relação à exposição e à resposta

Fonte: Hennekens & Buring (1987).

Um estudo observacional, entretanto, pode ser entendido genericamente como uma investigação empírica de tratamentos (ou exposições) e dos efeitos que eles causam, na qual, entretanto, diferentemente de um estudo experimental, o investigador não tem controle sobre a designação dos tratamentos considerados para as unidades. Isto é, as unidades não são designadas para os tratamentos por um dispositivo aleatório criado por um experimentador. Como já visto aqui, esse fato faz dos estudos observacionais mais sujeitos a vieses e conseqüentemente menos poderosos para avaliação de efeitos causais que os estudos experimentais. Assim, estudos observacionais são tipicamente empregados quando por alguma razão um estudo experimental não é possível, ou por questões éticas ou pela própria factibilidade do estudo. Assim, nos estudos observacionais, para compensar a insegurança quanto à comparabilidade dos grupos devida à falta de controle sobre o mecanismo de designação de tratamentos, o investigador deve se armar com uma teoria estatística pertinente. Essa teoria deve compreender, portanto, uma estrutura e um conjunto de ferramentas que forneçam medidas que considerem as evidências do estudo.

Procedimentos de ajustamento comumente utilizados tais como pareamento e estratificação, que procuram viabilizar a hipótese de homogeneidade de unidades e que podem, por exemplo, ser operacionalizados pelo escore de propensão, não garantem que os grupos tratado e controle sejam comparáveis sob todos os aspectos relevantes. Por exemplo, o pareamento garante que os dois grupos são comparáveis apenas quanto a algumas características previamente definidas. Se os grupos não são comparáveis antes da exposição aos tratamentos em razão de algum aspecto não considerado no pareamento, então uma eventual diferença na resposta observada pode ser apenas um reflexo dessa diferença inicial. O problema é de fato grave quando os grupos não são comparáveis e os dados disponíveis não são suficientes ou fracassam na tentativa de revelar esse fato. Técnicas

específicas, tais como a Análise de Sensibilidade, têm sido desenvolvidas para discutir essa questão. Se a estimativa de um efeito causal é insensível a variações plausíveis em supostos valores de co-variáveis não observadas, então uma interpretação causal se torna mais defensável. Assim, enquanto os procedimentos de ajuste comumente utilizados podem tentar corrigir vieses já conhecidos, os vieses ocultos, isto é, aqueles que necessitam de informações adicionais não observadas para serem visualizados, são tratados sob a técnica de Análise de Sensibilidade (Rosembaun, 1995).

Entre os diversos desenhos observacionais, um estudo de coorte ocupa uma posição privilegiada, sobretudo porque qualquer medida de risco pode ser obtida diretamente. Entretanto, a falta de controle do mecanismo de designação pode inviabilizar uma conclusão causal se não houver uma preocupação específica com os diversos vieses possíveis. No exemplo do uso de cafeína como fator causal para doença cardiovascular, viu-se que estimativas não ajustadas para uma medida de efeito pertinente deve ser confundida por muitas variáveis, tal como tabagismo, e deve também apresentar falta de validade de seguimento. Desse modo, o número de variáveis que deveriam ser controladas é muito grande para que um controle adequado fosse feito por estratificação. Assim, dado que a verdadeira dependência funcional dos riscos de doença cardiovascular para uso de cafeína e eventuais variáveis confundidoras é desconhecida, as estimativas obtidas por modelos multivariados provavelmente também estariam viesadas. E, mesmo que esse viés não fosse importante, as estimativas ainda permaneceriam confundidas em virtude da incapacidade de se medir acuradamente todas as variáveis envolvidas. Dado esse exemplo, no qual se constata que em um estudo de coorte devem existir diversas fontes de vieses de magnitudes desconhecidas e direções diferentes, qualquer conclusão de causalidade deve ser vista com muita cautela. No entanto, os elementos essenciais que constituem o Modelo de Rubin são facilmente identificados em um estudo de coorte. O desafio

seria, portanto, a identificação de um mecanismo ignorável, de um mecanismo que fosse dependente apenas dos dados observados. Essa não é obviamente uma tarefa simples, pois pressupõe necessariamente o estabelecimento de hipóteses adjacentes. E na possibilidade de tais hipóteses não serem passíveis de testes à luz dos dados observados, qualquer conclusão de causalidade deve vir acompanhada explicitamente da crença de que são válidas.

A situação é particularmente mais desfavorável quando o investigador está posicionado cronologicamente após a ocorrência da doença (Figura 2), quando a investigação causal é desenvolvida com base em um estudo de caso-controle ou de um estudo de coorte-retrospectivo. Assim, apesar da boa aderência dos estudos de coortes às condições do Modelo de Rubin, principalmente devido à forma de observação, isto é, buscando-se o efeito de causas postuladas e não as causas do efeito observado, as dificuldades práticas desses estudos têm estimulado um desenvolvimento teórico de desenhos de estudo retrospectivos, particularmente os estudos de caso-controle.

A característica básica que distingue um estudo de caso-controle é que a seleção das unidades é intencionalmente baseada na resposta dos indivíduos. Um grande complicador desse desenho é o fato de que a obtenção de medidas de efeito que estejam associadas explicitamente à noção de risco não é imediata, e depende da possibilidade de inclusão ou não dos casos no grupo dos controles selecionados. Aliás, a constituição de um grupo controle adequado é, sem dúvida, o grande desafio dos estudos e tem ocupado grande parte do tempo dos autores preocupados com tal questão. Segundo Greenland (1991), se os indivíduos que se tornam casos no período de risco considerado são inelegíveis para inclusão no grupo controle, como nos tradicionais desenhos de caso-controle, a hipótese de raridade da doença em questão será necessária para se estimarem riscos relativos com base em dados. Se, entretanto, tal como nos novos desenhos de caso-controle, esses indivíduos, mesmo sendo casos, são elegíveis para o grupo controle, então a hipótese de raridade pode ser dispensada.

As limitações do método caso-controle são bem conhecidas. Falta de validade devido a viés de seleção, mais freqüentemente causado por níveis altos de recusa à participação no estudo, gerando controles não representativos da população sob risco; erros de mensuração, particularmente devido a diferentes vieses de recordação entre casos e controles; e confundimento, a sempre presente possibilidade de que a associação encontrada é resultado de variáveis escondidas que falseiam uma associação causal, constituem as principais fragilidades desse estudo (Breslow, 1996). Entre suas vantagens, destacam-se as reduções tanto no tempo quanto no número de indivíduos necessários para atingir o mesmo poder estatístico que se atingiria em um estudo de coorte. Desse modo, adicionalmente às quatro condições de validade já apresentadas, deve-se acrescentar a condição de validade de seleção, tanto de casos quanto de controles. Haverá validade de seleção de casos quando o número de casos estudados – provavelmente menor que a população de casos, por alguma razão tal como a incapacidade de registrar todos –, que tenham ocorrido durante certo período, forneça estimativas não viesadas das prevalências, para a população de casos, das diferentes causas (níveis de exposição) consideradas. Validade de seleção de controles acontece sob a mesma lógica, substituindo casos por controles.

Com relação a procedimentos de ajuste, uma importante característica distingue os estudos de caso-controle e os de coorte. Nos estudos de coorte, pareamento refere-se à seleção das subcoortes de exposição, de modo a se forçar distribuições semelhantes entre as subcoortes para os fatores utilizados no pareamento. Esse procedimento deve prevenir confundimento para as variáveis envolvidas no pareamento, a menos que haja outra questão tal como uma falta de validade de seguimento relacionada a alguma variável do pareamento. Nos estudos de caso-controle, pareamento refere-se à seleção de indivíduos de maneira a forçar que as distribuições dos fatores considerados entre casos e controles sejam similares. Entretanto, sob esse

desenho, pareamento não previne confundimento. De fato, é sabido que um estudo de caso-controle pareado produz viés de seleção que pode ser, todavia, controlado pelo fator de pareamento. Um exemplo simples, presente em Greenland (1991), pode ser útil para esclarecer. Considere uma população sob risco, metade homem, em que os homens tendem a beber menos café que as mulheres e cerca de 75% dos casos de uma certa doença é de homens. Uma seleção de controles não viesada deveria ser constituída de 50% de homens. Se, entretanto, faz-se um pareamento dos controles aos casos pela variável sexo, cerca de 75% dos controles seriam constituídos de homens. E, visto que homens bebem menos café e seriam majoritários no grupo controle, a freqüência de usuários de cafeína estaria diminuída no grupo controle, subestimando a verdadeira proporção de usuários de cafeína na população sob risco. Como resultado, uma medida de efeito não ajustada seria superestimada. No entanto, as prevalências de usuários de cafeína no grupo controle específicas por sexo não são alteradas pelo pareamento. Logo, o viés de seleção poderia ser removido pelo ajustamento para a própria variável utilizada no pareamento. A conclusão é que pareamento pode exigir controle na análise para a variável utilizada no pareamento. Existe vasta literatura que discute os méritos desse procedimento; entre eles, a eventual redução na variância dos estimadores ajustados é considerado o principal. Assim, a fim de evitar um aumento no número de variáveis a serem controladas, o pareamento, quando indicado, deveria ser limitado apenas às variáveis que de fato fossem necessárias. Entretanto, a utilização já discutida do escore de propensão como instrumento para a construção de pares poderia ser uma excelente alternativa.

Do ponto de vista do Modelo de Rubin, a variável resposta para um estudo de caso-controle é a variável dicotômica que indica se a unidade é um caso ou um controle, isto é,

$$Y_S = \begin{cases} 1, & \text{caso a unidade seja um caso} \\ 0, & \text{caso a unidade seja um controle.} \end{cases}$$

E agora, diferentemente da situação ideal, deve-se buscar a causa desse efeito já observado. Nesse tipo de desenho, os parâmetros associados são as proporções populacionais (ou probabilidades), entre casos e controles, de indivíduos com e sem a exposição de interesse, tal como representadas na Tabela 7.

Como visto, os parâmetros de interesse em um estudo de causalidade são $\mathrm{E}(Y_t)$ e $\mathrm{E}(Y_c)$, que, quando se trata de variáveis dicotômicas, podem ser escritos como $\Pr(Y_t = 1)$ e $\Pr(Y_c = 1)$. Logo, as probabilidades retrospectivas da Tabela 7, genericamente dadas por $\Pr(S = k \mid Y_S = y)$, não apresentam nenhuma interpretação causal, pois sequer se referem ao evento correto. Entretanto, fazendo-se uso do teorema de Bayes, pode-se escrever

$$\Pr(Y_S = y \mid S = k) = \Pr(S = k \mid Y_S = y)\frac{\Pr(Y_S = y)}{\Pr(S = k)},$$

mas, $\Pr(Y_S = y \mid S = k) = \Pr(Y_k = y \mid S = k)$ e

$$\Pr(S = k) = \sum_y \Pr(S = k \mid Y_S = y)\Pr(Y_S = y).$$ Logo,

$$\Pr(Y_k = y \mid S = k) = \frac{\Pr(S = k \mid Y_S = y)\Pr(Y_S = y)}{\sum_y \Pr(S = k \mid Y_S = y)\Pr(Y_S = y)}.$$

Tabela 7 – Probabilidades populacionais retrospectivas

| Tratamentos ou causas | Resposta ($Y_S$) | |
|---|---|---|
| | Casos ($Y_S$=1) | Controles ($Y_S$=0) |
| $S = t$ | $\Pr(S=t \mid Y_S=1)$ | $\Pr(S=t \mid Y_S=0)$ |
| $S = c$ | $\Pr(S=c \mid Y_S=1)$ | $\Pr(S=c \mid Y_S=0)$ |
| TOTAL | 1 | 1 |

As probabilidades $\Pr(Y_k = y \mid S = k)$ são ditas 'prospectivas' porque os eventos condicionantes acontecem em algum tempo antes do evento de interesse. E são aquelas passíveis de ser obtidas diretamente por meio de estudos prospectivos, experimentais ou de coorte. A Tabela 8 ilustra essas probabilidades. A expressão anterior revela que as probabilidades prospectivas

podem ser determinadas com base nas probabilidades retrospectivas e nas proporções de casos e controles da população.

Com base nos parâmetros $\mathrm{E}(Y_t) = \Pr(Y_t = 1)$ e $\mathrm{E}(Y_c) = \Pr(Y_c = 1)$, viu-se que foi construída a medida de efeito causal conhecida como 'risco atribuível', definida como a diferença entre esses dois parâmetros. Entretanto, outras medidas de efeito de interesse podem ser definidas. O risco relativo, definido como

$$RR = \frac{\Pr(Y_t = 1)}{\Pr(Y_c = 1)}$$

e o *odds ratio*, cuja expressão correspondente é

$$OR = \frac{\Pr(Y_t = 1)}{\Pr(Y_t = 0)} \bigg/ \frac{\Pr(Y_c = 1)}{\Pr(Y_c = 0)} \quad ,$$

são largamente utilizados na literatura epidemiológica.

Tabela 8 – Probabilidades populacionais prospectivas

| Tratamentos ou causas | Resposta ($Y$) | | TOTAL |
|---|---|---|---|
| | Casos ($Y$=1) | Controles ($Y$=0) | |
| $S = t$ | $\Pr(Y_t=1 \mid S=t)$ | $\Pr(Y_t=0 \mid S=t)$ | 1 |
| $S = c$ | $\Pr(Y_c=1 \mid S=c)$ | $\Pr(Y_c=0 \mid S=c)$ | 1 |

Dentro de um estudo caso-controle tradicional em que a hipótese de raridade da doença não pareça razoável, a idéia de risco fica comprometida. Nesse caso, o parâmetro causal de interesse se limita ao *odds ratio*, que, com base na Tabela 7, pode ser estimado por

$$\hat{OR}_r = \frac{\Pr(S = t \mid Y_S = 1)}{\Pr(S = c \mid Y_S = 1)} \bigg/ \frac{\Pr(S = t \mid Y_S = 0)}{\Pr(S = c \mid Y_S = 0)}$$

Para verificar se essa medida pode ter uma interpretação causal, basta observar que ela se iguala à que seria obtida prospectivamente. Aplicando-se o teorema de Bayes às probabilidades que constituem $\hat{OR}_r$ e observando-se, como antes, que

$$\Pr(Y_S = y \mid S = k) = \Pr(Y_k = y \mid S = k),$$

chega-se à mesma expressão para o *odds ratio*, que pode ser obtida com base na Tabela 8, dada por

$$\hat{OR}_p = \frac{\Pr(Y_t = 1 \mid S = t)}{\Pr(Y_t = 0 \mid S = t)} \Big/ \frac{\Pr(Y_c = 1 \mid S = c)}{\Pr(Y_c = 0 \mid S = c)} = \hat{OR}_r$$

Apesar disso, ainda não é possível dizer se um particular estudo de caso-controle gozaria de uma interpretação causal. A expressão acima está para *OR* assim como

$$\hat{T} = \mathrm{E}(Y_t \mid S = t) - \mathrm{E}(Y_c \mid S = c) = \Pr(Y_t = 1 \mid S = t) - \Pr(Y_c = 1 \mid S = c)$$

está para

$$T = \mathrm{E}(Y_t) - \mathrm{E}(Y_c) = \Pr(Y_t = 1) - \Pr(Y_c = 1)$$

Logo, sem uma informação adicional não se pode falar sobre causalidade. No caso de $\hat{T}$, viu-se que sob randomização o problema estaria resolvido. Entretanto, um estudo de caso-controle nunca é randomizado. Assim, a alternativa consiste na observação de um conjunto de co-variáveis $X$ e na busca de um modelo para o mecanismo de designação que seja ignorável, de modo que, condicionalmente a $X$, $S$ e $Y$, sejam independentes.

Nossa finalidade, aqui, foi tentar inserir os desenhos tradicionalmente utilizados na pesquisa epidemiológica para atribuição de causa à lógica de causalidade presente no Modelo de Rubin. Se, em um nível mais geral, os estudos experimentais se contrapõem aos estudos observacionais pelo controle que o investigador possui sobre o mecanismo de designação de tratamentos, especificamente nos estudos observacionais, há também uma contraposição, porém muito menos significativa, dada pelos estudos de caso-controle e de coorte que se refere à noção apresentada de *timing* de observação. Assim, para o estabelecimento das condições de validade devido a Greenland (1991), essas questões foram consideradas, e, sob este aspecto, os estudos de coorte-retrospectivo se alinham aos estudos de caso-controle.

# Conclusão

Com base nas questões aqui apresentadas e discutidas, pode-se de forma sucinta tecer algumas considerações finais sobre a contribuição da estatística no processo de inferir causalidade, particularmente em epidemiologia, por meio da noção de causa contida no modelo de respostas potenciais devido a Rubin.

Dado que inferência causal e inferência estatística não se referem estritamente ao mesmo objeto, o primeiro ponto a ser considerado diz respeito à participação da estatística na discussão sobre causalidade. A partir dos trabalhos de Rubin, Holland e Dempster, entre os principais, não restam dúvidas quanto à relevância dos princípios estatísticos para inferência causal. Uma reflexão mais cuidadosa deve nos levar a concluir que não existiam dúvidas quanto a essa questão. Havia, sim, apenas um distanciamento, provavelmente em virtude de maior concentração na solução dos problemas mais característicos da estatística, ou seja, dos problemas associados à idéia de precisão ou modelagem. A questão da validade ficava a cargo da área de interesse, sendo o principal interesse em epidemiologia o julgamento da validade de hipóteses causais.

Outro sinal da importância da estatística em se preocupar também com o problema da causalidade pode ser extraído do comentário de Cox (1986) de que as questões surgidas explícita e implicitamente pelo artigo de Holland (1986) parecem ser mais importantes para os fundamentos da es-

tatística do que a discussão sobre a natureza da probabilidade. Assim, a formulação de um modelo estatístico para avaliação de causalidade não só amplia o campo de atuação da estatística, mas também outorga a ela o direito a uma co-participação no assunto principal – identificação de causas postuladas –, e não apenas, como outrora, um papel coadjuvante, de apoio às questões ligadas ao julgamento do acaso como explicação para os resultados. Gail (1996) parece respaldar essa conclusão ao afirmar que um estatístico que não esteja devidamente familiarizado com os princípios e as ferramentas estatísticas, que não conheça o método científico e que não visualize as vantagens e limitações tanto dos estudos experimentais quanto dos observacionais estará mal equipado para contribuir efetivamente para a solução de problemas complexos.

Ao mesmo tempo que a estatística reserva seu lugar na discussão de causa, também apresenta algumas condições que limitam sua participação em tal debate. Assim, uma definição epistemológica de causa não parece atrair o interesse da estatística. Epistemologia se refere à teoria da ciência, diferentemente da estatística, que se relaciona à tecnologia da ciência (Pereira, 1986). Entretanto, fazem-se necessários alguns conceitos, de modo a se viabilizar a identificação de causas. Com base na discussão apresentada, uma causa (ou tratamento) é entendida como qualquer estímulo ou conjunto de estímulos que conduza a uma resposta observável. E pode-se entender efeito causal como uma quantidade cuja variação em relação à causa alternativa (não exposição, por exemplo) deve ser atribuída à causa postulada (exposição). Desse modo, para fins de discussão sobre causalidade sob uma ótica estatística, substitui-se o conceito diverso de causa pelo conceito restrito de efeito causal, o qual parece ser mais apropriado em um contexto epidemiológico de acordo com a definição de causa de Rothman & Greenland (1998). E assim, complementando o que seja uma noção viável de causa dentro de uma perspectiva estatística, um atributo pessoal não

deve ser visto como uma causa. Isso não significa que um particular atributo não tenha um valor preditor, já que predição é simplesmente uma conseqüência de associação entre variáveis, que por sua vez não necessariamente envolve a noção de causa. Por sua vez, Granger (1969) formula sua noção de causa presente em modelos econométricos em torno da idéia de predição. Segundo ele, uma causa deve aumentar em um sentido probabilístico a habilidade de predizer o efeito.

Sobre a condição de que o caminho para se atribuir causalidade passa pela mensuração dos efeitos de causas postuladas e não pela busca da causa de um efeito observado, deve ser dito que tal limitação, apesar de revolucionária, não é nova e nem originariamente própria à estatística. Conta-se que Albert Einstein, quando professor da Politécnica de Zurique, causou verdadeiro escândalo entre seus colegas ao afirmar que o princípio básico de toda a ciência superior era priori-dedutivo e não posteriori-indutivo. Em outras palavras, o homem deve focalizar a 'causa' e daí partir para os 'efeitos' (Rohden, 1979). Nesse momento surgia então a magna pergunta: como atingir a causa, a não ser pelos efeitos? Talvez o Modelo de Rubin, com todas as suas limitações e propriedades explicitadas, seja uma possível resposta.

Cox (1992), apesar de reconhecer que a abordagem de causalidade contida no Modelo de Rubin com sua ênfase intervencionista representa uma noção mais específica de causa, aponta algumas limitações. A ausência explícita do entendimento do processo de causa adjacente aos dados envolvidos, isto é, o entendimento do mecanismo causal, é para ele uma séria limitação, já que este parece ser um importante aspecto da noção científica geral de causalidade. Um outro ponto levantado é o fato de que o modelo não torna claro o que está sendo mantido constante quando se varia hipoteticamente o agente causal. Em um ensaio clínico randomizado a análise comumente desenvolvida, conhecida como '*intention to treat*', avalia o

efeito de um tratamento comparado com outro onde a designação para o tratamento é seguida mesmo sob a possibilidade de outras variações, eventualmente presentes, que são permitidas pelo próprio desenho experimental. Por exemplo, se uma medicação suplementar não é controlada e é muito diferente entre os grupos sob os tratamentos considerados, então o efeito de um tratamento deveria incluir as conseqüências da medicação suplementar associada. Outro exemplo muito comum se refere à perfeita aderência de pacientes a uma terapêutica empregada. Normalmente os pacientes não seguem corretamente a posologia indicada. Este problema é conhecido como '*compliance*'. Assim, aquilo que se admite ser possivelmente diferente quando se modifica a designação ao tratamento é crucial para a interpretação de causalidade. Essas questões têm sido discutidas em um contexto epidemiológico (Lee et al., 1991; Imbens & Rubin, 1994; Efron & Feldman, 1991).

Um mecanismo causal é uma teoria científica que procura descrever os diversos processos biológicos, químicos, físicos e sociais pelos quais o tratamento produz seus efeitos. A complexidade por trás de um mecanismo causal parece bem representada pela máxima de Fisher quando argüido sobre o que poderia ser feito em estudos observacionais para elucidar o caminho que separa associação de causalidade: "Faça sua teoria elaborada" (apud Cochran, 1965). A elaboração a que Fisher faz alusão foi interpretada como a consideração, tanto quanto possível, de todas as conseqüências de uma hipótese causal estabelecida e como o planejamento detalhado de estudos para verificação de tais conseqüências. Entretanto, essa elaboração pode eventualmente ser simplificada pela construção de hipóteses causais mais específicas, de modo que as conseqüências dessas hipóteses possam ser consideradas e medidas. A máxima de Fisher é parafraseada por Rosembaun (1984), ao dizer que quando se estimam efeitos causais com base em um estudo observacional, é importante que se detalhe o 'mecanismo causal' tanto quanto o conhecimento científico vigente permita, verificando-se se os da-

dos do estudo em questão contradizem tal mecanismo. No entanto, deve-se observar que em experimentos randomizados efeitos causais podem ser estimados sem a especificação de um mecanismo causal. A necessidade de um mecanismo causal adjacente assume importância diferenciada em estudos observacionais e experimentais, principalmente porque, em experimentos randomizados, sabe-se que o mecanismo de designação de tratamentos é fortemente ignorável, enquanto em estudos observacionais esse mecanismo é apenas uma hipótese. Para Rosembaun (1984), um mecanismo causal em conjunto com um mecanismo de designação fortemente ignorável freqüentemente apresenta conseqüências testáveis. Diante desses pontos, pode-se concluir que a presença de um mecanismo causal adjacente ajuda na investigação de um dado fator causal. Entretanto, sua ausência não parece ser um impedimento.

A formulação do Modelo de Rubin é, sem dúvida, bastante elegante. Estudar causalidade com base na idéia de respostas potenciais torna bem clara a noção do que seja um efeito causal e, conseqüentemente, viabiliza a utilização do instrumental estatístico para sua avaliação. É claro que diversas condições precisam ser satisfeitas para que a formulação do modelo represente adequadamente a complexidade dos problemas reais. Quanto a sua estruturação, viu-se que a hipótese de estabilidade (ou hipótese de valor estável unidade-tratamento) era fundamental, apesar de ser possível estudar causalidade quando esta hipótese não for válida. Em epidemiologia, essa situação é bastante comum, devido à existência de doenças contagiosas em que a exposição à infecção (ou contágio) assume um papel-chave. Sob esse aspecto, Halloran & Struchiner (1995) concluem que a relação entre exposição à infecção, efeitos indiretos, confundimento e mecanismo de designação precisa de mais atenção. Níveis de exposição à infecção diferentes assumem função diferente quando se avaliam efeitos diretos e indiretos de um programa de intervenção, tal como vacinação. Quando se avaliam os efeitos diretos da vacinação sobre a suscetibilidade numa situação em que a expo-

sição à infecção não é a mesma entre os dois grupos (vacinado e não-vacinado), então a estimativa do efeito causal estará viesada.

A hipótese fundamental que permite inferências causais é que o vetor de respostas $(Y_t, Y_c)$ seja independente do mecanismo de designação. Sob randomização, viu-se que essa hipótese deveria ser satisfeita. Entretanto, diante de estudos observacionais apela-se para a observação de co-variáveis e para a hipótese de independência condicional, isto é,

$$(Y_t, Y_c) \perp S \mid X .$$

Como visto, um mecanismo que satisfaça a essa condição baseia-se na idéia de 'ignorabilidade' e deve gerar estimativas condicionalmente não viesadas para o efeito causal de interesse. É claro que a especificação de um mecanismo diferente de randomização que satisfaça à condição acima não deve ser trivial. Entretanto, estabelecido um mecanismo, os recursos que a estatística tradicionalmente utiliza para fazer inferências podem ser diretamente transferidos para o processo de inferência causal. Em particular, a abordagem bayesiana tem-se destacado no desenvolvimento de procedimentos e de teorias especificamente voltadas para a questão causal (Rubin, 1978; Schaffner, 1993; Gelman et al., 1995).

De forma bem genérica, pode-se dizer que se a estrutura representativa de causalidade presente no Modelo de Rubin é apropriada ou não, depende de uma identificação adequada dos elementos que a constituem. Assim, é usualmente fácil identificar as unidades $(U)$, os tratamentos $(K)$ e uma variável resposta $(Y)$ em estudos experimentais randomizados, pois nessa situação tudo está sob o controle do investigador e são estabelecidos *a priori*. Entretanto, estudos observacionais complexos freqüentemente proporcionam casos em que mesmo analistas experientes podem discordar sobre como identificar apropriadamente os elementos do modelo (Holland, 1986).

Entre as abordagens de que a epidemiologia tradicionalmente se tem utilizado para discutir causalidade, a mais objetiva parece ser aquela que se

baseia nos critérios estabelecidos por Hill (1965). A abordagem filosófica e o Modelo de Rothman são de difícil operacionalização. No entanto, apesar de os critérios de Hill ajudarem a determinar se uma exposição é fator de risco para uma dada doença, sua aplicação também é limitada. Como visto, apenas o critério de temporalidade é necessário e nenhum dos outros é suficiente para uma interpretação causal. Adesão a qualquer um deles sem demais considerações poderia resultar em falsas conclusões. Por exemplo, é muito importante considerar a precisão e a validade interna dos resultados de cada estudo.

Os estudos epidemiológicos podem perfeitamente ser inseridos na lógica de causalidade devida a Rubin. Um ensaio clínico, por já ser randomizado, tem uma inserção imediata. Os estudos observacionais necessitariam da especificação de um mecanismo de designação de tratamentos que resgatasse (ou inferisse) a condição de exposição que cada indivíduo já possuía no momento do início do estudo. É nesse ponto que as contribuições da epidemiologia e da estatística devem se combinar para constituir uma análise mais abrangente e se aproximar mais da difícil tarefa que é o estabelecimento de causa. Se de um lado a estatística pode contribuir com o Modelo de Rubin, métodos de inferência ou modelagem, a epidemiologia pode contribuir com a fundamental discussão de validade das estimativas. Nesse sentido, a classificação de validade devida a Greenland (1991), apesar de ainda não formalizada, não só estende a discussão de viés tradicionalmente contida nos estudos epidemiológicos por meio da tríade 'confundimento, viés de informação e viés de seleção', como também permite uma visualização mais adequada das limitações do Modelo de Rubin. Como visto, esse modelo é fortemente dependente do estabelecimento de um mecanismo de designação de tratamentos. Logo, uma especificação não apropriada para esse mecanismo implicaria falta de validade de especificação e, conseqüentemente, uma possível interpretação de causa ficaria comprometida. Por outro lado, já foi dito que um mecanismo de designação incorreto estaria relacionado a

confundimento, ou seja, à falta de validade de comparação. Na realidade, de acordo com os conceitos contidos na classificação de Greenland, o que acontece de fato, é uma falta de validade de especificação. Uma eventual confusão pode surgir porque qualquer violação das hipóteses adjacentes, tanto por erro de especificação quanto por erro de mensuração ou mesmo por falta de validade de seguimento, implica impossibilidade de comparação imediata entre os grupos sob os tratamentos considerados.

Há confundimento, isto é, falta validade de comparação, sempre que existir diferença entre os riscos de doença entre os grupos sob comparação, independentemente do fator causal em questão. Em outras palavras, haverá confundimento sempre que o grupo de indivíduos não expostos não representar o que aconteceria com os indivíduos expostos, caso eles não tivessem sido expostos. Planejar um estudo real em que tal conceito possa ser operacionalizado não parece ser tarefa simples. Entretanto, pode-se depositar uma razoável credibilidade nessa hipótese se o mecanismo é randomizado, já que sob esta condição espera-se que as distribuições de eventuais variáveis confundidoras não sejam muito diferentes nos dois grupos de exposição. E, como já dito aqui, a probabilidade de confundimento grave pode ser feita tão pequena quanto necessária aumentando-se o tamanho dos grupos randomizados. Além disso, sob randomização, qualquer eventual confundimento que ainda persista após ajustamentos adequados será considerado pelo erro padrão da estimativa, dado que a especificação do modelo estatístico utilizado para computar o efeito estimado e seu erro padrão esteja correta (Greenland, 1991).

Assim, atribuição de causa é dependente do estabelecimento de hipóteses que, de acordo com o estudo em questão, podem ser ou não plausíveis. As hipóteses de causa transiente, estabilidade temporal, independência, homogeneidade de unidades, efeito constante e valor estável unidade-tratamento aqui discutidas, combinadas ou não, constituem as premissas neces-

sárias para se estabelecer um modelo teórico de causalidade. Acrescente-se a elas ainda as questões relativas à operacionalidade peculiar a cada estudo, tais como viés de informação, viés de seleção, erros de mensuração, dados censurados e erros na especificação do modelo probabilístico de estimação, cuja consideração é fundamental para o estudo de causa. No caso de doenças contagiosas, exigem-se hipóteses especificamente sobre a exposição à infecção. A validade de uma atribuição de causa é, portanto, dependente de todas essas condições, com a agravante de em muitas situações não ser possível colocá-las à prova. Sendo assim, o modelo estatístico de causalidade devido a Rubin não deve ser visto como a panacéia da questão causal. Tal como a abordagem filosófica, ou os critérios estabelecidos por Hill ou ainda o modelo de causas componente/suficiente de Rothman, o Modelo de Rubin apresenta condicionantes que muitas vezes não permitem uma atribuição segura de causa para determinado fator. Sua contribuição sem o devido acompanhamento desses condicionantes, apresentados explicitamente, é de pouca valia.

Paralelamente, na atualidade têm sido sugeridas outras abordagens de causalidade, tal como os diagramas causais (Pearl, 1995), ou a incorporação de novos elementos ao Modelo de Rubin, tal como o de variáveis instrumentais (Angrist et al., 1996). Discute-se, ainda, a união da estrutura de causalidade devida a Rubin às abordagens filosófica e de modelos causais gráficos (Schaffner, 1993).

Concluindo, o julgamento final sobre os efeitos causais de um tratamento freqüentemente dependerá do acúmulo de evidências obtidas por meio de uma série de estudos. A contribuição de cada estudo depende da capacidade do investigador, ao interpretar os resultados de um estudo que mostre uma associação consistente com uma hipótese causal, de listar e discutir todas as explicações alternativas, incluindo hipóteses diferentes ou possíveis vieses.

# Referências Bibliográficas

Angrist, J. D.; Imbens, G. W. & Rubin, D. B. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444-472, 1996.

Breslow, N. E. Statistics in epidemiology: the case-control study. *Journal of the American Statistical Association*, 91:14-28, 1996.

Cochran, W. G. The planning of observational studies of human populations. *Journal of the Royal Statistical Society*, series A, 128:234-265, 1965.

Cox, D. R. *Planning of Experiments*. Nova York: John Wiley & Sons, 1958.

Cox, D. R. Comment on 'statistics and causal inference'. *Journal of the American Statistical Association*, 81:963-964, 1986.

Cox, D. R. Causality: some statistical aspects. *Journal of the Royal Statistical Society*, series A, 155:291-301, 1992.

D'Agostino Jr., R. B. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17(19):2265-2282, 1998.

Dawid, A. P. Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association*, 95:407-448, 2000.

Dempster, A. P. Causality and statistics. *Journal of Statistical Planning and Inference*, 25:261-278, 1990.

Efron, B. & Feldman, D. Compliance as an explanatory variable. *Journal of the American Statistical Association*, 86:9-17, 1991.

Evans, A. S. Causation and disease: a chronological journey. *American Journal of Epidemiology,* 108(4):249-258, 1978.

Gail, M. H. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In: Moolgavkar, S. H. & Prentice, R. L.

(Eds.) *Modern Statistical Methods in Chronic Disease Epidemiology*. Nova York: Wiley, 1986.

GAIL, M. H. Statistics in action. *Journal of the American Statistical Association*, 91:1-13, 1996.

GELMAN, A. et al. *Bayesian Data Analysis*. Londres: Chapman & Hall, 1995.

GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424-438, 1969.

GREENLAND, S. Limitations of the logistic analysis of epidemiologic Data. *American Journal of Epidemiology*, 110(6):693-698, 1979.

GREENLAND, S. Reader reactions: confounding in epidemiologic studies. *Biometrics*, 45:1309-1310, 1989a.

GREENLAND, S. Commentary: modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*, 79(3):340-349, 1989b.

GREENLAND, S. Randomization, statistics and causal inference. *Epidemiology*, 1(6):421-429, 1990.

GREENLAND, S. Concepts of validity in epidemiological research. In: HOLLAND, W. W., DETELS, R. & KNOX, G. (Eds.) *Oxford Textbook of Public Health*. 2.ed. Methods of Public Health. Nova York: Oxford University Press, 1991. v.2.

GREENLAND, S. Invited commentary on 'causes'. *American Journal of Epidemiology*, 141(2):89, 1995.

GREENLAND, S. Causal Analysis in the health sciences. *Journal of the American Statistical Association*, 95:286-289, 2000.

GREENLAND, S. & ROBINS, J. M. Identifiability, exchangeability, and epidemiological Confounding. *International Journal of Epidemiolgy*, 15(3):413-419, 1986.

GREENLAND, S., ROBINS, J. M. & PEARL, J. Confounding and collapsibility in causal inference. *Statistical Science*, 14(1):29-46, 1999.

HALLORAN, M. E. & STRUCHINER, C. J. Causal inference in infectious diseases. *Epidemiology*, 6:142-151, 1995.

HENNEKENS, C. H. & BURING, J. E. *Epidemiology in Medicine*. Boston: Little, Brown and Company, 1987.

HILL, A. B. The environment and disease: association or causation? *Proceeding of the Royal Society of Medicine*, 58:295-300, 1965.

HOLLAND, P. W. Statistics and causal inference. *Journal of the American Statistical Association*, 81:945-970, 1986.

HOLLAND, P. W. Reader reactions: confounding in epidemiologic studies. *Biometrics*, 45:1310-1316, 1989.

HOLLAND, P. W. & RUBIN, D. B. Causal inference in retrospective studies. *Evaluation Review*, 12(3):203-231, 1988.

IMBENS, G. W. & RUBIN, D. B. *Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance*. Boston: Harvard University, Depto. of Economics, 1994.

KEMPTHORNE, O. Logical, epistemological and statistical aspects of nature-nurture data interpretation. *Biometrics*, 34:1-23, 1978.

KLEINBAUM, D. G.; KUPPER, L. L. & MORGENSTERN, H. *Epidemiologic Research: principles and quantitative methods*. Nova York: Van Nostrand Reinhold, 1982.

LEE, Y. J. et al. Analysis of clinical trials by treatment actually received: is it really an option? *Statistics in Medicine*, 10:1595-1605, 1991.

LITTLE, R. J. & RUBIN, D. B. Causal effects in clinical and epidemiological studies via potencial outcomes: concepts and analytical approaches. *Annual Review of Public Health*, 21:121-145, 2000.

MIGON, H. S. & GAMERMAN, D. *Statistical Inference: an integrated approach*. Londres: Arnold, 1999.

PEARL, J. Causal diagrams for empirical research. *Biometrika*, 82(4):669-710, 1995.

PEREIRA, B. B. *Estatística: a tecnologia da ciência*. Rio de Janeiro: Instituto de Matemática da UFRJ, 1986. (Estudos e Comunicações, 27).

ROBINS, J. & GREENLAND, S. The role of model selection in causal inference from nonexperimental data. *American Journal of Epidemiology*, 123(3):392-402, 1986.

ROHDEN, H. *Einstein: o enigma da matemática*. 3.ed. São Paulo: Alvorada, 1979.

ROSENBAUM, P. R. From association to causation in observation studies: the role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association*, 79:41-48, 1984.

ROSENBAUM, P. R. Coherence in observational studies. *Biometrics*, 50(2):368-374, 1994.

ROSENBAUM, P. R. *Observational Studies*. Nova York: Springer-Verlag, 1995. (Springer Series in Statistics).

ROSENBAUM, P. R. & RUBIN, D. B. the central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41-55, 1983.

ROSEMBAUN, P. R. & RUBIN, D. B. Comment: estimating the effects caused by treatments. *Journal of the American Association*, 79:26-28, 1984.

ROTHMAN, K. J. Causes. *American Journal of Epidemiology*, 104(6):587-592, 1976.

ROTHMAN, K. J. *Causal Inference*. Chestnut Hill, MA: Epidemiology Resources Inc., 1988.

ROTHMAN, K. J. & GREENLAND, S. *Modern Epidemiology*. 2.ed. Filadélfia: Lippincott Williams & Wilkins, 1998.

RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688-701, 1974.

RUBIN, D. B. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, 6(1):34-58, 1978.

RUBIN, D. B. Discussion of 'randomization analysis of experimental data: the Fisher Randomization Test' by D. Basu. *Journal of the American Statistical Association*, 75:591-593, 1980.

RUBIN, D. B. Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25:279-292, 1990a.

RUBIN, D. B. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5:472-480, 1990b.

RUBIN, D. B. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 47:1213-1234, 1991.

SCHAFFNER, K. F. Clinical trials and causation: bayesian perspectives. *Statistics in Medicine*, 12:1477-1494, 1993.

SIEMIATYCKI, J. & THOMAS, D. C. Biological models and statistical interactions: an example from multistage carcinogenesis. *International Journal of Epidemiology*, 10(4):383-387, 1981.

SMITH, T. M. F. & SUGDEN, R. A. Sampling and assignment mechanisms in experiments, surveys and observational studies. *International Statistical Review*, 56(2):165-180, 1988.

STRUCHINER, C. J. & HALLORAN, M. E. Randomization and baseline transmission in vaccine field trials. *Rollins School of Public Health of Emory University, Department of Biostatistics*, 1996. (Technical Report, 96-97).

SUSSER, M. Judgment and causal inference: criteria in epidemiologic studies. *American Journal of Epidemiology*, 105(1):1-15, 1977.

VANDENBROUCKE, J. P. Should we abandon statistical modeling altogether? *American Journal of Epidemiology*, 126(1):10-13, 1987.

WAINER, H. Adjusting for differential base rates: lord's paradox again. *Psychological Bulletin*, 109(1):147-151, 1991.

WICKRAMARATNE, P. J. & HOLFORD, T. R. Confounding in epidemiologic studies: the adequacy of the control group as a measure of confounding. *Biometrics*, 43:751-765, 1987.

WINKELSTEIN JR., W. Invited commentary on 'Judgment and causal inference: criteria in epidemiologic studies'. *American Journal of Epidemiology*, 141(8):699-700, 1995.

YERUSHALMY, J. & PALMER, C. E. On the methodology of investigations of etiologic factors in chronic diseases. *Journal of Chronical Diseases*, 10:27-40, 1959.